

# **MODIFIED TECHNIQUES FOR MODELLING SURVIVAL DATA WITH COMPETING RISKS**

**BY**

**DAUDA, KAZEEM ADESINA**

**B.Sc., M.Sc. Statistics (Ilorin)**

**06/55EG048**

**July, 2017**

# **MODIFIED TECHNIQUES FOR MODELLING SURVIVAL DATA WITH COMPETING RISKS**

**BY**

**DAUDA, KAZEEM ADESINA**

**B.Sc., M.Sc. Statistics (Ilorin)**

**06/55EG048**

**A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS,  
UNIVERSITY OF ILORIN, ILORIN FOR THE AWARD OF  
DOCTOR OF PHILOSOPHY (Ph.D.) IN STATISTICS.**

## **CERTIFICATION**

This is to certify that this thesis is an original work carried out by DAUDA, KAZEEM ADESINA with matriculation number 06/55EG048 in the Department of Statistics, Faculty of Physical Science, University of Ilorin. The thesis has been read and approved as meeting the requirements of the award of Doctor of Philosophy in Statistics of the University of Ilorin, Ilorin, Nigeria.

-----  
**Dr. W. B. Yahya**  
**(Supervisor)**

-----  
**Date**

-----  
**Dr. O. O. M. Sanni**  
**(Departmental Postgraduate Coordinator)**

-----  
**Date**

-----  
**Prof. B. L. Adeleke**  
**(Head of Departmental & Chief Examiner)**

-----  
**Date**

-----  
**External Examiner**

-----  
**Date**

## **DEDICATION**

This project is dedicated to my precious and forbearing wife and daughter, Hajia Sofiyat Omolara Dauda and Aisha Anike Dauda for their support and patience throughout the course of this work.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my supervisor Dr. W. B. Yahya (Associate Professor of Statistics) for his ideas, guidance, support, and encouragement and excellent supervision until the completion of this project. If not for his constructive criticisms, valuable suggestions, and fatherly advice that I received, this project would not have seen the light of the day. May Almighty Allah provide him and his family their needs in this earth and hereafter (amen).

My sincere appreciation and thanks also go to all Professors in the Department of Statistics, University of Ilorin, Nigeria and all other lecturers in the Department for their support, guidance and encouragement. The unalloyed support, encouragement and prayers of Dr. M. K. Garba is highly appreciated.

I immensely acknowledge the presence of my esteemed parents and siblings, especially Mrs. Morufat Ademola and Hammed Akorede Dauda, for their fervent prayers and support over several decades all of which have assisted me through the thin and thick of my academic endeavor.

I will forever be grateful to my dearest and caring wife and daughter Hajia Sofiyat Omolara Dauda and Aishat Anike Dauda for their steadfastness, patience, prayers and support throughout the course of this study.

Part of this research study was supported by Kwara State University (KWASU) through the Tertiary Education Training Fund (TETFUND). I enormously thank this body and KWASU for the financial support.

I equally appreciate all the lecturers in the Department of Statistics and Mathematical Sciences, KWASU, starting from the Head of Department, Prof. Bayo Lawal to the last faculty.

I am also grateful to my friends for their helpful comments and suggestion through the course of this program: Waheed Alabi Banjoko, Kabir Opeyemi Olorede, Abdulwasii Ibrahim Opeyemi, Kabir Opeyemi Otun, Suraju Adesina Adepoju and others too numerous to mention. Finally, all praise, glory and adoration are due to Allah who in His infinite mercy has spared my life to the successful completion of the Ph.D. Programme at the University of Ilorin.

## ABSTRACT

A major drawback of traditional Cox Proportional Hazard (CPH) technique for modeling survival data is its inability to identify and classify subjects based on their inherent biological associations with respective risk endpoints as well as its inability to model the possible interaction effects among the covariates. To this end, this work aims at developing some techniques for modeling survival time data, especially with competing risks to address these challenges. The objectives of this study are to: (i) develop survival tree-based technique via the Classification and Regression Tree (CART) using within-node homogeneity and between-node heterogeneity methods; (ii) extend the Multi-Layer Perceptron Artificial Neural Network (MLPANN) for modeling survival data with competing risks; (iii) extend the Multivariate Adaptive Regression Splines (MARS) method to model classical survival data using Cox-Snell residuals; and (iv) compare the efficiencies of the proposed methods with the existing methods using simulated and real life data sets.

The survival trees were formed by calculating the reduction in *impurity* going from the *parent node* to the *child nodes* via an impurity function. Two impurity functions of Deviance and Cox-Snell Residuals were used for both the sum of squares and absolute value impurity functions in within-node homogeneity. The Cox-Snell Residual was employed as a common response in MLPANN and MARS models. For the between-node heterogeneity trees, different inference procedures for testing the equality of two cumulative incidence functions were employed. The efficiency of the various methods were assessed using root mean square, mean absolute and relative square errors.

Findings of the study were that:

- (i) the within-node homogeneity and between-node heterogeneity survival Trees with competing risks were developed to model the data and classify the subjects based on their associated risks;
- (ii) a MLPANN-based method for modeling survival data with competing risks was developed;
- (iii) method for modeling classical survival data using MARS techniques was developed and was found capable of modeling the possible interactions among the covariates;
- (iv) the proposed survival trees model determined the cutpoint of the tree better than the existing methods;
- (v) the proposed MLPANN-based method was shown to be more efficient than the Cox PH model using three real life data sets; and
- (vi) results obtained based on simulated and real-life HIV data sets showed that the proposed survival MARS model was better than the existing methods considered.

This study concluded that the proposed methods were suitable for identifying subjects with their associated risks as well as being able to model possible interaction effects among the covariates on survival time. They are therefore recommended whenever interest is focused on identifying subjects with their biological risks through prognostic variable biomarkers for diagnostic purposes.

# TABLE OF CONTENTS

|                               |             |
|-------------------------------|-------------|
| <i>Title Page</i> .....       | <i>i</i>    |
| <i>Certification</i> .....    | <i>ii</i>   |
| <i>Dedication</i> .....       | <i>iii</i>  |
| <i>Acknowledgements</i> ..... | <i>iv</i>   |
| <i>Abstract</i> .....         | <i>v</i>    |
| <i>List of Tables</i> .....   | <i>ixii</i> |
| <i>List of Figures</i> .....  | <i>xii</i>  |

## **CHAPTER ONE .....1**

### **BACKGROUND OF THE STUDY .....1**

|                                                           |    |
|-----------------------------------------------------------|----|
| 1.0 Introduction.....                                     | 1  |
| 1.1 Survival Time Data.....                               | 2  |
| 1.2 Survival Analysis with Competing Risk Modelling.....  | 3  |
| 1.3 Cox Proportional Hazard Model .....                   | 5  |
| 1.3.1 Model diagnostic for the Cox PH model.....          | 6  |
| 1.4 Cause-Specific Hazard Model.....                      | 8  |
| 1.5 Data Mining in Statistics.....                        | 8  |
| 1.5.1 Classification and Regression Tree (CART) .....     | 9  |
| 1.5.2 Artificial Neural Network.....                      | 9  |
| 1.5.3 Multivariate Adaptive Regression Spline (MARS)..... | 13 |
| 1.6 Statement of the Problem.....                         | 14 |
| 1.7 Research Motivation.....                              | 14 |
| 1.8 Aim and Objectives of the Study .....                 | 14 |
| 1.9 Research Questions .....                              | 15 |
| 1.10 Definition of Basic terms .....                      | 16 |
| 1.11 Outline of the thesis.....                           | 18 |

## **CHAPTER TWO .....19**

### **LITERATURE REVIEW .....19**

|                                                                                                    |           |
|----------------------------------------------------------------------------------------------------|-----------|
| 2.1 Data Mining .....                                                                              | 19        |
| 2.2 Data Mining Models .....                                                                       | 19        |
| 2.2.1 <i>Decision Trees</i> .....                                                                  | 20        |
| 2.2.2 <i>Neural Networks</i> .....                                                                 | 21        |
| 2.2.3 <i>Multivariate Adaptive Regression Splines (MARS)</i> .....                                 | 24        |
| 2.3 Related Research in Survival Tree-based with and without Competing Risks .....                 | 26        |
| 2.4 Related Research in Multivariate Adaptive Regression Spline Survival Analysis Based .....      | 29        |
| 2.5 Related Research on Artificial Neural Network in Survival Analysis .....                       | 31        |
| 2.5 Summary and Research Gaps .....                                                                | 33        |
| <b>CHAPTER THREE .....</b>                                                                         | <b>35</b> |
| <b>RESEARCH METHODOLOGY .....</b>                                                                  | <b>35</b> |
| 3.1 Classification and Regression Tree (CART) Methodology .....                                    | 36        |
| 3.1.1 <i>General Algorithm to Grow Tree</i> .....                                                  | 37        |
| 3.1.2 <i>Algorithm to Prune Tree</i> .....                                                         | 38        |
| 3.1.3 <i>Task 1: Survival Tree-Based via within-node homogeneity</i> .....                         | 39        |
| 3.1.4 <i>Task 2: Survival Tree-Based via between-node heterogeneity</i> .....                      | 50        |
| 3.2 Task 3: Extension of Artificial Neural Network in Survival Analysis with Competing Risks ..... | 61        |
| 3.2.1 <i>Relative Important by Garson Algorithm</i> .....                                          | 64        |
| 3.3 Task 4: Extension of Multivariate Adaptive Regression Splines in Survival Analysis ..          | 65        |
| 3.3.1 <i>Simulation Studies of the Survival MARS</i> .....                                         | 68        |
| 3.4 Summary of Chapter Three .....                                                                 | 71        |
| <b>CHAPTER FOUR .....</b>                                                                          | <b>72</b> |
| <b>DISCUSSION OF RESULTS .....</b>                                                                 | <b>72</b> |
| 4.1 <i>Survival Tree-Based via within-node homogeneity (Task 1) Numerical Results</i> .....        | 72        |
| 4.1.1 <i>Cutpoint detection results on survival based-tree via within-node homogeneity</i> .....   | 73        |

|                                                                                                   |            |
|---------------------------------------------------------------------------------------------------|------------|
| 4.1.2 Data Structure Performance results on survival based-tree via within-node homogeneity ..... | 82         |
| 4.1.3 Application of survival based-tree via within-node homogeneity on Real-life Dataset.....    | 84         |
| 4.2 Survival Tree-Based via Between-Node Heterogeneity (Task 2) Numerical Results.....            | 98         |
| 4.2.1 Application of survival based-tree via between-node heterogeneity on Real-life Dataset..... | 106        |
| 4.2.2 Estimation of Optimal cutpoint on between-node heterogeneity.....                           | 108        |
| 4.3 Results of Artificial Neural Network in Survival Analysis with Competing Risks (Task 3) ..... | 108        |
| 4.3.1 Comparing the Proposed Model with Classical Competing Risk Model.....                       | 110        |
| 4.2.3 Implementation CMLPANN on Double-Blinded Randomized Clinical Data ..                        | 112        |
| 4.4 Results of Multivariate Adaptive Regression Splines in Survival Analysis (Task 4)...          | 116        |
| 4.4.1 Results of the Simulation on Survival MARS.....                                             | 116        |
| 4.4.2 Application of Survival MARS to Pneumocystis Cariniip Neumonia Data.....                    | 119        |
| <b>CHAPTER FIVE.....</b>                                                                          | <b>122</b> |
| <b>Summary, Conclusion and Recommendation .....</b>                                               | <b>122</b> |
| 5.1 Summary of Results .....                                                                      | 122        |
| 5.2 Conclusion .....                                                                              | 127        |
| 5.3 Recommendations .....                                                                         | 128        |
| 5.4 Suggestions for Future Studies.....                                                           | 129        |
| <i>Reference .....</i>                                                                            | <i>130</i> |
| <i>Appendices.....</i>                                                                            | <i>140</i> |

## LIST OF TABLES

|                                                                                                                                       |    |
|---------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 3. 1: Splitting rule criteria on sum-of-squares impurity functions.....                                                         | 42 |
| Table 3. 2: Splitting rule criteria on Absolute value impurity functions.....                                                         | 43 |
| Table 3. 3: Model for Data Structure Detection via Exponential Distribution.....                                                      | 47 |
| Table 3. 4: Model for Data Structure Detection via Lognormal Distribution .....                                                       | 48 |
| Table 3. 5: Description of models used for simulation in data structure performance .....                                             | 50 |
| Table 3. 6: Simulation Procedures of Classification Tree Based on Between Node Heterogeneity<br>.....                                 | 61 |
| Table 3. 7: Statistics for generating survival times across the specified distributions.....                                          | 69 |
| Table 4.1: Low censoring rate 25% .....                                                                                               | 73 |
| Table 4.2: Moderate censoring rate 50%.....                                                                                           | 74 |
| Table 4.3: High censoring rate 75%.....                                                                                               | 74 |
| Table 4.4: Low censoring rate 25% .....                                                                                               | 79 |
| Table 4.5: Moderate censoring rate 50%.....                                                                                           | 79 |
| Table 4.6: High censoring rate 75%.....                                                                                               | 80 |
| Table 4.7: Examining tree structure based on 1 event of interest, through exponential<br>distribution. N=500 .....                    | 82 |
| Table 4.8: Examining tree structure with one event-of-interest, through log-normal distribution.<br>N=500 .....                       | 83 |
| Table 4.9: Results of bone marrow transplant through cause 1 (relapse) .....                                                          | 85 |
| Table 4.10: Results of bone marrow transplant through cause 2 (death in remission) .....                                              | 89 |
| Table 4.11: The Summary Statistics of Age, Weight index (WTINDEX) and Serum<br>Haemoglobin gm/100ml (SERUMHG) in prostate cancer..... | 94 |
| Table 4.12: Cutpoint selected based on the terminal nodes using measure of impurities (event 1).<br>.....                             | 94 |

|                                                                                                                                                                                                                                                      |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 4.13: Simulation with Exponential Distribution over 50% censoring rate and 0.5 cut-points .....                                                                                                                                                | 99  |
| Table 4.14: Relative performance of the proposed and the existing methods across the sample sizes .....                                                                                                                                              | 100 |
| Table 4.15: Performance of methods based on system time (Exponential Distribution).....                                                                                                                                                              | 102 |
| Table 4.16: Simulation with Log-normal Distribution over 50% censoring rate and 0.5 cut-points .....                                                                                                                                                 | 103 |
| Table 4.17: Relative performance of the proposed and the existing methods across the sample sizes .....                                                                                                                                              | 105 |
| Table 4.18: Performance of methods based on system time (Lognormal Distribution) .....                                                                                                                                                               | 106 |
| Table 4.19: Summary Statistics of covariates Age and hemoglobin (g/l).....                                                                                                                                                                           | 107 |
| Table 4.20: Maximum cutpoint generated across the selected test.....                                                                                                                                                                                 | 108 |
| Table 4.21: Double-blinded randomized clinical trial with first and second events (death due to Prostatic Cancer and Other diseases) .....                                                                                                           | 110 |
| Table 4.22: Melanoma skin cancer incidence with first and second events (death from melanoma and other causes).....                                                                                                                                  | 111 |
| Table 4.23: Mayo Clinic Primary Biliary Cirrhosis Data with first and second events (death from Transplant and Remission) .....                                                                                                                      | 112 |
| Table 4.24: Resilient back propagation with weight for event1 (death due to Prostatic Cancer).....                                                                                                                                                   | 113 |
| Table 4.25: Variable Important by Garson Algorithm (1991).....                                                                                                                                                                                       | 115 |
| Table 4.26: Results of Survival MARS with Exponential distributed survival time. The amount of non-linear effects of covariates $x_1$ (in %) and its interactions with $x_2$ with Martingale and Cox-Snell residuals as responses were reported..... | 117 |
| Table 4.27: Results of the simulation studies on Weibull distributed ( $v = 2, \lambda = 5$ ) survival time with Martingale and Cox-Snell residuals. ....                                                                                            | 117 |
| Table 4.28: Results of the simulation studies on Gompertz distributed survival time with Martingale and Cox-Snell residuals. ....                                                                                                                    | 118 |

|                                                                                                                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 4.29: Results of Survival MARS models using Cox-Snell and Martingale residuals on real life (PCP) data. Results of the classical Cox PH model fitted to these data were also reported in the table. .... | 120 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

## LIST OF FIGURES

|                                                                                                                                                                 |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Fig. 1.1 : Schematic presentation of a two-state model for survival data with the transition intensity $\alpha t$ from state 0 to state 1. ....                 | 2  |
| Fig. 1. 2: Competing risks model for mortality from K causes .....                                                                                              | 4  |
| Fig. 1. 3: Structure of Biological Neuron.....                                                                                                                  | 11 |
| Fig. 1. 4: Artificial Neuron Design .....                                                                                                                       | 12 |
| Fig. 2.1: Sample decision tree-Partial view .....                                                                                                               | 21 |
| Fig. 2.2: A neural network structure .....                                                                                                                      | 22 |
| Fig. 2.3: Artificial Neuron configurations, with bias as additional Input .....                                                                                 | 23 |
| Fig. 3.1: Framework for Classification and Regression Tree (CART) research methodology .....                                                                    | 35 |
| Fig. 4. 1: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 25% censoring rate (Exponential Distribution) ..... | 76 |
| Fig. 4.2: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 50% censoring rate (Exponential Distribution) .....  | 76 |
| Fig. 4.3: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 75% censoring rate (Exponential Distribution) .....  | 77 |
| Fig. 4.4: Boxplot of exponential distribution .....                                                                                                             | 77 |
| Fig. 4.5: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 25% censoring rate (Lognormal Distribution).....     | 81 |
| Fig. 4.6: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 50% censoring rate (Lognormal Distribution).....     | 81 |
| Fig. 4.7: The plot of the various cutpoint against the Root Mean Square Errors and Mean Absolute Errors at 75% censoring rate (Lognormal Distribution).....     | 82 |

|                                                                                                                                                            |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Fig. 4.8: Within node tree for data from bone marrow transplant through cause 1 (relapse) via martingale residual .....                                    | 86 |
| Fig. 4.9: Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residual.....              | 87 |
| Fig. 4.10: Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual.....                                      | 87 |
| Fig. 4.11: Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residual.....             | 88 |
| Fig. 4. 12: Within node tree for data from bone marrow transplant through cause 1 (relapse) via cox-snell residual .....                                   | 88 |
| Fig. 4.13: Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via Cox-Snell residual.....              | 89 |
| Fig. 4.14: Within node tree for data from bone marrow transplant through cause 2 (death in remission) via martingale residual .....                        | 90 |
| Fig. 4.15: Graphical representation of the variable important in bone marrow transplant through cause 1 (relapse) via martingale residual.....             | 90 |
| Fig. 4.16: Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual .....                                     | 91 |
| Fig. 4.17: Graphical representation of the variable important in bone marrow transplant through cause 2 (death in remission) via martingale residual. .... | 92 |
| Fig. 4.18: Within node tree for data from bone marrow transplant through cause 2 (death in remission) via Cox-Snell residual .....                         | 93 |
| Fig. 4.19: Graphical representation of the variable important in bone marrow transplant through cause 2 (death in remission) via Cox-Snell residual .....  | 93 |
| Fig. 4.20: Plot of Martingale residual against the patient Age using existing prognostics classification tree.....                                         | 96 |

|                                                                                                                                          |     |
|------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 4.21: Plot of Deviance residual against the patient Age using proposed prognostics<br>classification tree.....                      | 97  |
| Fig. 4.22: Plot of Cox-Snell residual against the patient Age using proposed prognostics<br>classification tree.....                     | 97  |
| Fig. 4.23: Histogram of estimated cutpoint in exponential distribution .....                                                             | 100 |
| Fig. 4. 24: Histogram of estimated cutpoint in Log-normal distribution .....                                                             | 105 |
| Fig. 4.25: Plot of a trained neural network including trained synaptic weights and basic<br>information about the training process. .... | 114 |
| Fig. 4.26: Relative importance of the eight covariates for response (probability of survival)<br>variable using CMLP Model.....          | 115 |

# CHAPTER ONE

## BACKGROUND OF THE STUDY

### 1.0 Introduction

The most common methods for analyzing survival data with and without competing risks are the Cox Proportional Hazard (Cox-PH) methods (Cox, 1972) and the Cumulative Incidence Function (CIF) techniques (Gray, 1988) among others. However, the major drawbacks of these excellent classical techniques lie first in the difficulty encountered in interpreting their results in situations where both categorical and metrical covariates are present in the models. Another limitation is their inability to model possible (complex) interactions that might exist between these covariates. The third limitation of these methods is their inability to group subjects/samples into sub-groups based on their different levels of risk.

In recognition of all the limitations of the classical modelling techniques for event history data as highlighted above and many more, it therefore becomes quite imperative to evolve alternative techniques to model survival data with competing risks endpoint in the presence of categorical and metrical covariates all of which might describe the biological characteristics of the subjects/patients as well as the associated prognostic variables depending on the kind of ailment being managed.

To achieve the above tasks, this work seeks to explore and extend the techniques of a number of robust machine learning methods for modelling survival data that are characterized with the aforementioned challenges. More specifically, two groups of nonparametric methodologies for investigating competing risk modeling based on recursive partitioning via the Classification and

Regression Tree (CART) (Fiona, 2008 and Jiawei et al., 2012) and the one based on Multivariate Adaptive Regression Spline (MARS) for censored survival data are discussed in this chapter. Also, the concepts of Artificial Neural Network (ANN) method, a data mining technique like CART and MARS, was briefly discussed here while the adaptations of various methods for modelling survival data with competing risks were provided in details in the subsequent chapters. Finally, the traditional techniques of the Cox PH and Cause Specific Hazards (CSH) for analyzing both survival and competing risks models were equally discussed briefly.

## 1.1 Survival Time Data

Multi-state models described how individuals move between a finite number of states. The simplest example of multi-state model is the survival model with one transient state ‘0: alive’ and one absorbing state “1: dead”. This model is characterized by the distribution of the survival time  $T$ , and this indicates the time from a particular origin (time = 0) to the existence of the event of interest such as “death”. The distribution of time could be characterized by the distribution  $F(t) = Prob(T \leq t)$  or consistently by the Survival function  $S(t) = 1 - F(t) = Prob(T > t)$ . If every subjects (objects) were supposed to be in state 0 at time 0, then  $F(t)$  is also the transition probability from state 0 to state 1 for the time interval ranges between 0 to  $t$ . The distribution of time can also be characterized by the hazard function  $\alpha(t) = -\frac{d(\log S(t))}{dt}$  (Cox, 1972).

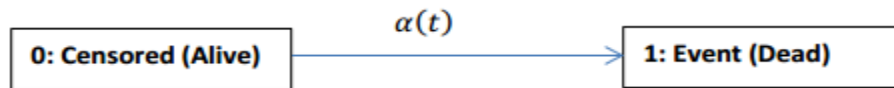


Fig. 1.1 : Schematic presentation of a two-state model for survival data with the transition intensity  $\alpha(t)$  from state 0 to state 1.

Thus,  $\alpha(\cdot)$  is the transition intensity from state 0 to state 1. This scenario is illustrated by Fig 1.1.

In Fig 1.1 there are two states 0 and 1, with only one possible transition from state 0 to 1. The

conforming transition intensity  $\alpha_{01}(t)$  is given by the hazard rate function  $\alpha(t)$ , while  $\alpha_{10}(t) = 0$  for all  $t$ , that is, state 1 is an absorbing state. The initial distribution is degenerate in 0 with  $\pi_0(0) = 1$  and the process is Markovian (Andersen, 1993). In a nutshell, survival time data typically means time until the occurrence of particular events. The *time* (period) may be years, months, weeks, days or hours from the beginning of the follow-up of a subject/object until an event of interest occurs. Here, event may refer to death, disease incidence, relapse from remission, return to work, or specified experience of interest that may happen to any subject/object at the end of follow-up period.

One of the major concepts in survival analysis is **censoring**, and it occurs when we have information about the subjects/objects survival time, but the exact survival time is unknown. Generally, censoring occurs as a result of the following: when subject/object does not experience the event of interest till the end of the study; when subject/object is lost due to follow-up during the study; or when the subject/object withdraws from the study due to one reason or another. There are three types of censoring which include: **right censoring**, **left censoring** and **interval censoring**. Right-censoring occurs when the true survival time is equal to or greater than observed survival time; left-censoring occurs when the true survival period is less than or equal to observed survival time, while interval-censoring occurs when the true time is equal to observed survival time (David and Moeschberger, 1978).

## 1.2 Survival Analysis with Competing Risk Modelling

The historical background of competing risks can be traced back to 1760 when Daniel Bernoulli analyzed the risks of death from smallpox compared to death due to other causes (David and Moeschberger, 1978). Up to date, the Bernoulli research work has laid the foundation for the

theoretical background of the competing risks. In a simpler form, the situation when subject can experience two or more none overlapping events at a specific time is regarded as competing risks modelling (Pintilie, 2006 and Crowder, 2012, 2001).

Furthermore, this model (competing risks modelling) has one transient state ‘0: alive’ and a number, k of absorbing states, state , 1, h, h=1,2,..., k corresponding to ‘death from cause h’. Thus, there are  $p=k+1$  states. The model is illustrated for  $k^{\text{th}}$  causes in Fig 1.2.

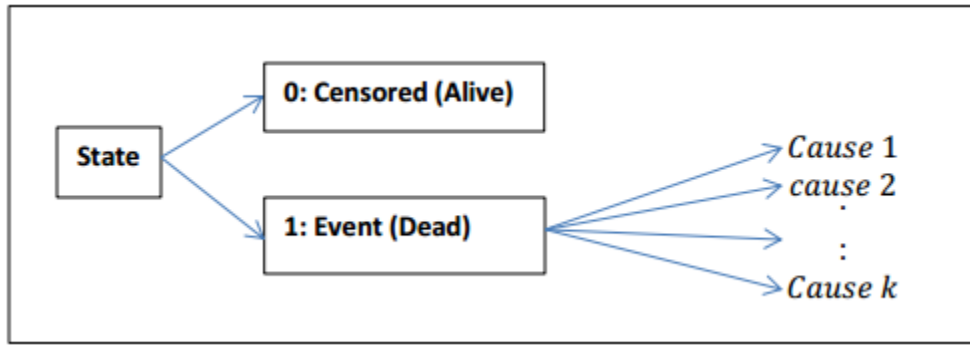


Fig. 1. 2: Competing risks model for mortality from k causes

The transition intensities  $\alpha_{0h}(t)$  for  $h=1,\dots,k$  is given by the cause-specific hazard functions:

$$\alpha_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\text{prob}(\text{dead from cause } h \text{ by time } t+\Delta t / T \geq t)}{\Delta t},$$

where T is the survival time. The initial distribution is degenerate in 0 with only one transient state of the model, i.e.,

$\alpha_{kj}(t) = 0$  for all  $h \neq 0$  and all j. The transition probabilities are given by the survival function

$$P_{00}(o, t) = S(t) = \text{Prob}(T > t) = \text{EXP}(-\int_0^t \sum_h^k \alpha_h(u) du),$$

and the so-called cumulative incidence functions

$$P_{0h}(o, t) = \int_0^t S(u -) \alpha_h(u) du, \quad h = 1, \dots, k \quad (1.1)$$

There are several techniques for regression analysis in the face of competing risks. Cause Specific Hazard, Latent Failure, Pseudo-Observation, Sub-distribution hazard and Cumulative Incidence function (see equation 1.1). However, two popular and robust techniques for analyzing competing risks are Cause Specific-Hazards (CSHs) and Cumulative Incidence Function proposed by Fine and Gray (1999). All these techniques have their own drawbacks, which will be explained in the subsequent pages. However, various researchers have shown that the Fine and Gray (1999) methods uses Cumulative Incidence Functions directly, and there is less mix-up in interpreting the predictor effects when translating from cause-specific hazards than in the Cox model (Fiona, 2008).

### 1.3 Cox Proportional Hazard Model

The Cox PH regression model (Cox 1972) assumes that the covariates multiplicatively shift the baseline hazard function and asserts that the hazard rate for the  $j^{th}$  subject in the data is

$$h(t/X) = h_0(t)exp(\beta X) \quad (1.2)$$

where  $X$  is the vector of covariates (predictors) and  $\beta$  is the vector of regression coefficients that are to be estimated from the data (Cleves et al., 2004). The baseline hazard ( $h_0(t)$ ) represents the hazard of an ‘average’ subject of the data, thus the hazard of a particular subject is just a multiple of the baseline hazard. In the Cox model, the most important assumption is the proportionality of the hazard rate. That is, for any given input factor  $X_j$  with two binary levels (0,1), the hazard rate (HR) is given by;

$$HR = \frac{h(t/X)}{h(t/X=0)} = \frac{h_0(t)exp(\beta X)}{h_0(t)} = exp(\beta X) \quad (1.3)$$

independent of time  $t$ . This means that the hazard rate of a particular subject remains the same over the whole observation time. That is, the influence of all the covariates in the model is independent of the survival/observational time  $t$  (Kriner, 2007; Yahya and Ulm, 2009).

### 1.3.1 Model diagnostic for the Cox PH model

Residuals for survival data are slightly different from other types of models, due to the censoring.

Before discussion of

residuals, we need an important basic information below:

If  $T_i$  (the survival time for the  $i^{\text{th}}$  individual) has survivorship function  $S_i(t)$ , then the transformed random variable  $S_i(T_i)$  (i.e., the survival function evaluated at the actual survival time  $T_i$ ) should be from a uniform distribution on  $[0,1]$ , and hence  $-\log(S_i(T_i))$  (i.e. transformed random variable) should be from a unit exponential distribution.

#### i. **Generalized (Cox-Snell) Residuals (CSR)**

The implication of the last result in equation (1.2) is that if the model is correct, the estimated cumulative hazard for each individual at the time of their death or censoring should be like a censored sample from a unit exponential. This quantity is called the generalized or Cox-Snell residual. Here is how the generalized residual might be used. Suppose we fit a CPH model in equation (1.2), then the cumulative hazard and survival estimates of the CPH model is given in equation (1.4) and (1.5) respectively (Klein and Moeschberger. (2003) as follows:

$$\hat{S}(t, X) = [\hat{S}_0(t)]^{\text{Exp}(\beta X)} \quad (1.4)$$

and

$$\hat{h}(t, X) = -\log(\hat{S}(t, X)) \quad (1.5)$$

Thus, the Cox-Snell (Cs) residuals can be estimated through equation (1.5) as follows:

$$\widehat{CS} = \delta_i - \hat{h}(t, X_j), \quad (1.6)$$

where  $\delta_i$  is the events (status).

ii. **Martingale Residuals (MR)**

Martingale residuals are used to examine the functional form of a given covariate in order to make the best use of it to explain its influence on the Cox proportional hazard model. If the test tells us that we could not introduce a covariate into the model linearly, we might need to transform the covariate  $X$  into  $\log(X)$ ,  $\exp(X)$ ,  $X\log X$  or even a discretized version of the covariate.

Suppose we have a vector  $X(t)$  of possible time-dependent covariates. Let  $N_j(t)$  have a value equals to 1 at time  $t$  if this individual has experienced the event of interest and 0 if the individual is yet to experience the event of interest. Let  $Y_j(t)$  be the indicator that individual  $j$  is under study at a time just prior to time  $t$ . Finally, let  $b$  be a vector of regression coefficients and  $\hat{H}_0(t)$  the Breslow estimator of the cumulative baseline hazard rate (Breslow, 1974), then the martingale residual is defined as:

$$\hat{M}_j = N_j(\infty) - \int_0^\infty Y_j(t) \exp[b'X_j(t)] d\hat{H}_0(t), j = 1, \dots, n \quad (1.7)$$

iii. **Deviance Residuals (DR)**

Martingale residuals are defined on  $(-\infty, 1]$  and thus highly skewed. To determine whether this skewness does in any way affect the outcome, a new method was applied to the transformation of the martingale residuals, named deviance residuals, which are symmetrically defined on  $(-\infty, \infty)$ .

If all covariates are time-fixed, the deviance residual for subject  $j$  is defined as

$$\hat{D}_j = \text{sign}(\hat{M}_j) \sqrt{(-\hat{M}_j - N_j \log(\frac{N_j - \hat{M}_j}{N_j})), j = 1, \dots, n}$$

(1.8)

(Klein et al., 2003).

#### 1.4 Cause-Specific Hazard Model

A cause-specific hazard is the instantaneous risk of failure from a specific cause given that failure (from any cause) has yet to happen. It is mathematically defined as:

$$\lambda(t, x) = \log_{\Delta t \rightarrow 0} \frac{Pr\{t \leq T < t + \Delta t | T \geq t, x\}}{\Delta t}$$

(1.9)

The cause-specific hazard rate that represents the instantaneous risk of dying from cause  $h$  is

$$\lambda_h(t, x) = \log_{\Delta t \rightarrow 0} \frac{Pr\{t \leq T < t + \Delta t, h = k | T \geq t, x\}}{\Delta t}$$

(1.10)

This is the conditional probability that a subject with covariates  $x$  dies in the interval  $[t, t + \Delta t)$  and the cause of death is the  $h^{th}$  cause, given that the subject was alive just before time  $t$ . We turn the probability into a rate dividing by  $\Delta t$  and then take the limit as  $\Delta t \rightarrow 0$  (Prentice et al., 1978).

#### 1.5 Data Mining in Statistics

The process of discovering interesting knowledge, which includes: patterns, associations, changes and significant structures from large amount of data set from statistical perspective is regarded as *Data Mining* (Fayyad, 1996). It is also seen as nontrivial extraction of implicit, previously unknown, and potentially useful information from data using machine learning, statistical and

visualization techniques (Piatetsky and Frawley, 1991). Data mining has proven to be valuable in numerous applications, due to increase in technology around the globe (Vapnik, 1998; Hastie et al., 2001; Witten and Frank, 2005; Felici and Vercellis, 2007). The majority of the problems studied in data mining community can be categorized as: Decision Trees, Neural Networks, Rule Induction, Nearest Neighbors, and Genetic Algorithms (Hastie et al., 2001).

### **1.5.1 Classification and Regression Tree (CART)**

Classification and Regression Tree (CART) is the most commonly and popularly used tool for exploring large data building. Moreover, it is one of the most stretchy, instinctive and powerful information analytic and can be applied in the area of determining prognostic subgroups with the same results within each subgroup but different outcomes between the subgroup (Fiona, 2008 and Jiawei et al., 2012). It categorizes the successive classification structure that recursively partition the set of progressive observation into smaller subgroups based on binary division of the covariates (Ibrahim et al., 2008).

CART was first brought together by Breiman et al. (1984) and it is gaining more attention recently because of its high predictive ability in model building (tree-building) for difficult data mining problems. There are other number of techniques that can handle this difficulties in machine learning, they includes: CHAID, MARS, Artificial Neural Network (Yahya et al. 2012, Yahya, 2014) and so on.

### **1.5.2 Artificial Neural Network**

A mathematical ideal that tries to simulate from the operational and functionalities of biological neural network is known as Artificial Neural Network (ANN) (Andrej et. al., 2011). A neural network consists of an interconnected group of artificial neurons, and it processes information

using a connectionist approach for computation. They are powerful tools for modelling linear and non-linear functional relationship between the covariates and response variable, especially when the underlying assumptions that govern the classical model fail (Frauke and Stefan, 2010). The basic building block of every artificial neural network is artificial neuron, that is, a simple mathematical model (function). Such a model has three simple sets of rules: multiplication, summation and activation. At the entrance of artificial neuron the inputs are weighted, this means that every input value is multiplied with individual weight. In the middle section of artificial neuron is sum function that sums all weighted inputs and bias (Andrej et. al., 2011).

A **neuron** is a nervous cell, which is the basic functional building element of the nervous system. Only the human cortex consists of approximately 13 to 15 billion neurons which are arranged into a hierarchical structure of six different layers. Moreover, each neuron can be connected with about 5000 other neurons. The neurons are autonomous cells that are specialized in transmission, processing, and storage of information, which are essential for the realization of vital functions of the organism. The structure of a neuron is schematically depicted in fig. 1.3 (Sima, 1998).

#### i. Modelling of Biological Neuron: Neurodes

The neuron is formed by transferring the signal received by the *soma*, which also has both input and output transfer channels, i. e. the *dendrites* and the *axon*, respectively. The axon is branched out into many '*terminals*', which are terminated by a membrane to contact the **thorns** of dendrites of other neurons as it is depicted in Figure 1.3. A **synapse** serves as a unique inter-neuron interface to transfer the information. The degree of synaptic permeability bears all important knowledge during the whole life of the organism. From the functional point of view, the synapses are classified in two types: the **excitatory** synapses which enable impulses in the nervous system

to be spread and the **inhibitory** ones, which cause their attenuation. A memory trace in the nervous system probably arises by encoding the synaptic bindings on the way between the receptor and effector (Sima, 1998 and Gurney , 1997). The mathematical model of neuron shall be given in the methodology of this research.

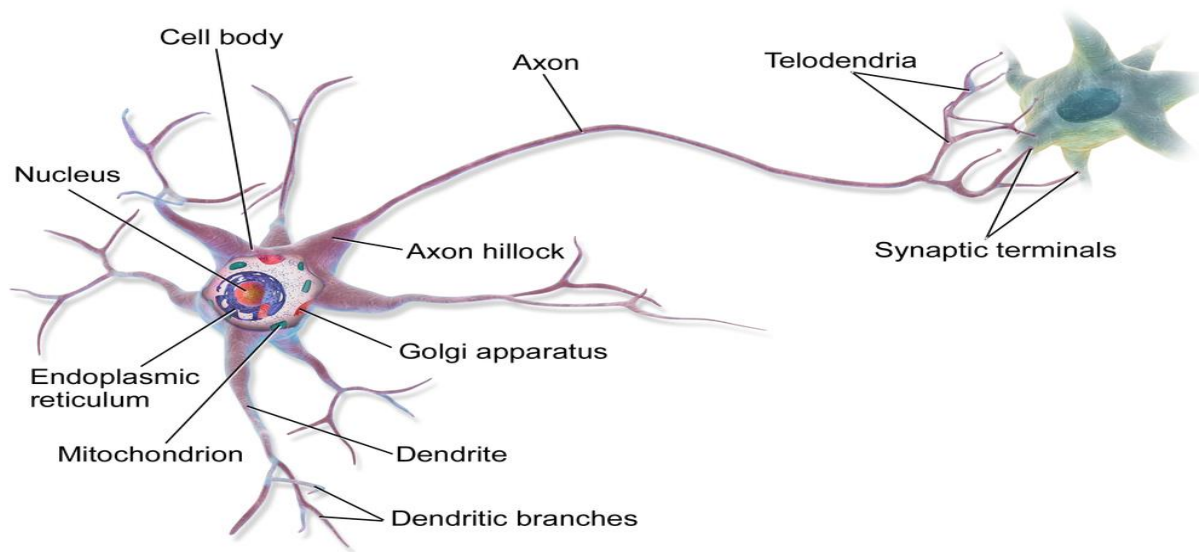


Fig. 1. 3: Structure of Biological Neuron

(Source:[https://en.m.wikipedia.org/wiki/Neuron#/media/File%3ABlausen\\_0657\\_MultipolarNeuron.png](https://en.m.wikipedia.org/wiki/Neuron#/media/File%3ABlausen_0657_MultipolarNeuron.png))

Simulation through Electrical input can occur in neural cell and may not fire its action potentially (Wickens et al., 1988). When the neural network is formed, it can encode information to mimic the human neuron; this process is called neural coding. The information can be encoded and can also be decoded. The decoding process is sometimes referred to as recalling process and the encoding process is called training or learning. This makes the artificial neural network efficient and reliable.

In case of biological neuron (see Fig.1.3) information comes into the neuron via dendrite, soma processes the information and passes it on via axon. In case of artificial neuron (see figure 1.4) the

information comes into the body of an artificial neuron via inputs that are weighted (each input can be individually multiplied with a weight). The body of an artificial neuron then sums the weighted inputs, bias and “processes” the sum with a transfer function. At the end, an artificial neuron passes the processed information via output(s) (Andrej et. al., 2011 and Krenker et al., 2009). Mathematical representation of this model will be given in the subsequent pages.

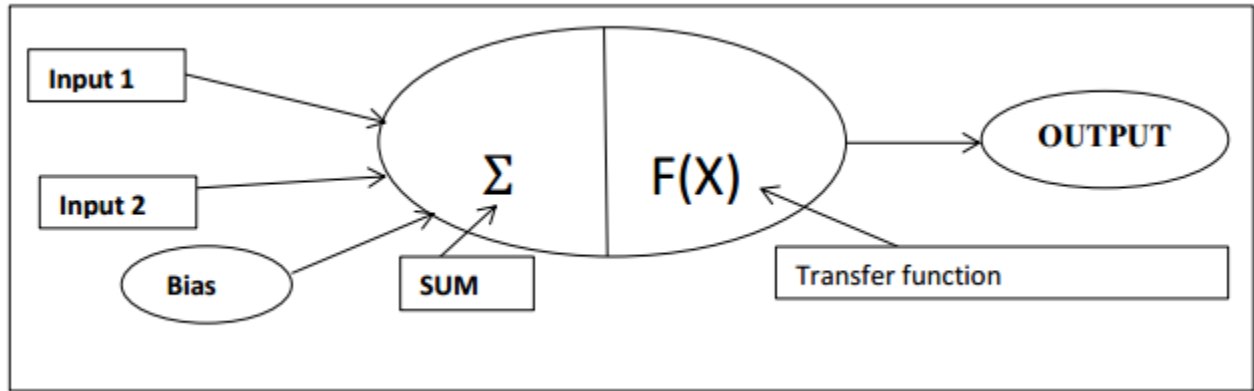


Fig. 1. 4: Artificial Neuron Design

## ii. Learning process

Among the most important properties of neural networks is its ability to learn from the environment and it uses this process to improve the performance of the output variable. It does this process through the grid iterative methods by adjusting and applying it to the synaptic weight and threshold of neuron. In general, learning is a process by which the free parameters of a neural network are adapted through a continuing process of stimulation by the environment in which the network is embedded (Hajek, 2005 and Jain and Mao, 1994). Mathematically, let  $w_{kj}(n)$  denote the synaptic weight value  $w_{kj}$  at time  $n$ . At time  $n$  an adjustment  $\Delta w_{kj}(n)$  is applied to the synaptic  $w_{kj}(n)$ , yielding the updated value

$$w_{kj}(n + 1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (1.11)$$

There is no unique learning algorithm for the design of neural networks. But there exists a “kit tools” represented by a diverse variety of learning algorithms. These learning algorithms differ based on their advantages and the way they adjust  $\Delta w_{kj}$  to the synaptic weight  $w_{kj}$  to be formulated. On that note, the learning algorithm includes: Error correction learning, Hebbian learning, Supervised learning and unsupervised learning. For full details of these learning see Hajek, 2005.

### 1.5.3 Multivariate Adaptive Regression Spline (MARS)

Multivariate adaptive regression splines (MARS) are used to identify the linearity and nonlinearity effects and interactions between two or more covariates in classical regression models (Breiman et al., 1984). This approach has been embraced to the area of survival by few researchers. The use of standard analysis of covariate effects such as Cox proportional hazard function can be misleading and difficult to interpret if we have both continuous and binary predictors in the model at the same time. Due to the limitations of these standard methods of analyzing survival models, the MARS technique was embraced as an alternative means of analyzing survival data (Kriner, 2007).

In contrast to the global parametric modeling technique of Linear Regression Model (LRM), the MARS technique developed by Friedman (1991) is a nonparametric modeling approach of the form  $y = f(x) + \epsilon$ , where  $y$  is a response variable,  $x = (x_1, x_2, \dots, x_p)^T$ , is a vector of  $p$  predictors and  $\epsilon$  is an additive stochastic component which is assumed to have zero mean and finite variance (Weber et al., 2012).

In recent years, MARS has been successfully applied to many areas of science and technology. These include predicting object-oriented software maintainability by Zhou et al. (2007), waste water treatment by Tsai (2005) and species population from presence of data by Elith et al. (2007).

It also has applications in speech modeling, credit scoring, breast cancer pattern determination, and examining the impact of information technology investment on productivity (Weber et al., 2012).

Among the recent applications of MARS to survival analysis was that of Kriner (2007) in which MARS was applied to Survival Analysis using both martingale and deviance residuals.

## **1.6 Statement of the Problem**

Chemotherapy, radiation, and/or surgery are vital treatments for cancer disease. Survival analysis, which is analysis of survival time after a patient had received any of these clinical treatments has been the primary method for evaluating the effectiveness of such treatment. It is therefore important to evaluate the patient's survival in an optimal manner and to identify the risk factors that are really associated to the prognostic risks. There are numerous methods for analyzing survival problems which include: Cox Proportional Hazard (CPH), Cause-specific hazard and Cumulative Incidence Function (CIF), but all of them involve a number of restrictive assumptions which often makes their implementations very cumbersome. Not only this, results that arise from the use of these classical methods are difficult to interpret in clinical sense. Therefore, there is need to adopt methods that can circumvent these problems.

## **1.7 Research Motivation**

The most common methods for analyzing survival data with and without competing risks end points are the classical CPH and CIF. These models are difficult to interpret if we have both continuous and categorical predictors in the models. Also, these models lacked ability to handle the interactions that might exist between these crop of predictors and incapability to provide a clear method for grouping subject/objects into group based on their levels of risk. This study therefore aimed to provide an alternative robust machine leaning technique that would be capable to

circumvent the above stated challenges and be useful to efficiently model survival data with or without competing risks.

### **1.8 Aim and Objectives of the Study**

The aim of this study is to undertake the efficiency measurement of some statistical methods for modelling survival data with and without competing risks. However, the specific objectives of this study are to;

- i. develop survival tree-based techniques via the Classification and Regression Tree (CART) using within-node homogeneity and between-node heterogeneity methods;
- ii. extend the Multi-Layer Perceptron Artificial Neural Network (MLPANN) for modeling survival data with competing risks;
- iii. extend the Multivariate Adaptive Regression Splines (MARS) method to model classical survival data using Cox-Snell residuals; and
- iv. compare the efficiencies of the proposed methods with the existing methods using simulated and real-life data sets.

### **1.9 Research Questions**

In view of the above objectives, the following research questions are addressed:

- i. Can survival tree-based methods be developed to model survival data within the context conjectured in this study?
- ii. Is it possible to extend Multi-Layer Perceptron Artificial Neural Network (MLPANN) to survival analysis with competing risks?

- iii. Can Multivariate Adaptive Regression Splines (MARS) method be extended to survival analysis?
- iv. Can data mining-based survival models yield results that can be clinically and/or operationally meaningful?

### 1.10 Definition of Basic terms

A partial listing of symbols and other basic terms used frequently in this thesis:

| Symbols/Notations/Functions | Symbols/Notations/Functions Descriptions         |
|-----------------------------|--------------------------------------------------|
| CHS                         | Cox Proportional Hazard                          |
| CSHs                        | Cause Specific Hazards                           |
| CART                        | Classification and Regression Tree               |
| ANN                         | Artificial Neural Network                        |
| MPLANN                      | Multi-layer Perceptron Artificial Neural Network |
| MARS                        | Multivariate Adaptive Regression Splines         |
| CSR                         | Cox-Snell Residual                               |
| MR                          | Martingale Residual                              |
| DR                          | Deviance Residual                                |
| CIF                         | Cumulative Incidence Function                    |
| GCV                         | Generalized Cross-Validation                     |
| c                           | Cutoff/Cutpoint/Threshold value                  |
| $h$                         | Node                                             |
| CSRSS                       | Cox-Snell Residual Sum-of-Square                 |
| CSRAV                       | Cox-Snell Residual Absolute Value                |
| DRSS                        | Deviance Residual Sum-of-Square                  |
| DRAV                        | Deviance Residual Absolute Value                 |
| MRSS                        | Martingale Residual Sum-of-Square                |
| MRAV                        | Martingale Residual Absolute Value               |
| MAE                         | Mean Absolute Error                              |

|                            |                                                                              |
|----------------------------|------------------------------------------------------------------------------|
| RMSE                       | Root Mean Square Error                                                       |
| RSE                        | Relative Square Error                                                        |
| MAE <sub>1</sub>           | Mean Absolute Error for event 1                                              |
| MAE <sub>2</sub>           | Mean Absolute Error for event 2                                              |
| $ T^t $                    | Expected Terminal nodes                                                      |
| $T^{NS}$                   | Neyman smooth Test Statistic                                                 |
| $T^{KS}$                   | Kolmogorov Smirnov Test Statistic                                            |
| $T^{PM}$                   | Pepe Moris Test Statistic                                                    |
| $T^G$                      | Gray Test Statistic                                                          |
| $T^{FG}$                   | Fine and Gray Test Statistic                                                 |
| $T^{LR}$                   | Log-rank Test Statistic                                                      |
| $o(x)$                     | Response variable in ANN                                                     |
| RI                         | Relative Important                                                           |
| Covariate/Input            | Explanatory variable                                                         |
| Outputs                    | Predicted values                                                             |
| Targets or Training values | Dependent variables                                                          |
| $f(t)$                     | Probability Distribution Function                                            |
| $F(t)$                     | Cumulative Distribution Function                                             |
| $S(t)$                     | Survival Function                                                            |
| $h(t)$                     | Hazard Function                                                              |
| $h(t, x)$                  | Cox Proportional Hazard Model function                                       |
| $h_0(t)$                   | Baseline hazard function                                                     |
| $H(t)$                     | Cumulative Hazard function                                                   |
| $R(t)$                     | Reliability function                                                         |
| $\hat{h}(t)$               | Estimated Hazard function                                                    |
| $T_i$                      | Survival Time for $i^{\text{th}}$ individual                                 |
| $I(x \leq k)$              | An indicator function whose value is 1 if $x \leq k$ , 0 otherwise           |
| $\delta_i$                 | Status variable whose value is 1 if individual $i$ had the event of interest |
| $Y_i(t)$                   | The outcome variable whose value is 1 if subject $i$ is under risk at        |

|                        |                                                                   |
|------------------------|-------------------------------------------------------------------|
|                        | time $t$                                                          |
| $N_i(t)$               | Counting process: $(N_i(t) \in \{0,1\})$                          |
| $\mathcal{F}_i$        | Filtration or history                                             |
| $\emptyset(h)$         | Impurity Function for node $h$                                    |
| $\emptyset_{SS}(h)$    | Sum-of-square Impurity Function for node $h$                      |
| $\emptyset_{AV}(h)$    | Absolute Value Impurity Function for node $h$                     |
| $\emptyset_{CSRSS}(h)$ | Cox-Snell Residual Sum-of-Square Impurity Function for node $h$   |
| $\emptyset_{CSRAV}(h)$ | Cox-Snell Residual Absolute Value Impurity Function for node $h$  |
| $\emptyset_{DRSS}(h)$  | Deviance Residual Sum-of-Square Impurity Function for node $h$    |
| $\emptyset_{DRAV}(h)$  | Deviance Residual Absolute Value Impurity Function for node $h$   |
| $\emptyset_{MRSS}(h)$  | Martingale Residual Sum-of-Square Impurity Function for node $h$  |
| $\emptyset_{MRAV}(h)$  | Martingale Residual Absolute Value Impurity Function for node $h$ |
| $\emptyset_\alpha(h)$  | Cost Complexity Function                                          |
| $\hat{r}_{CS}^m$       | Neuron Output via Cox-Snell residual (Observed output)            |
| $\hat{r}_{CS}^{*m}$    | Predicted Neuron Output via Cox-Snell residual                    |

### 1.11 Outline of The Thesis

The remaining parts of this thesis are organized as follows. The review of literatures that are relevant to this study are provided in Chapter Two while Chapter Three presents the research methodology. Chapter Four presents the various results from the study and provides detail discussion of the results. Finally, summary of the results from the study, conclusions, contribution to knowledge and suggestions for future studies are provided in Chapter Five.

## **CHAPTER TWO**

### **LITERATURE REVIEW**

This current section discusses the different data mining techniques and Modelling methods that are well known in the literature; it also distinguishes their main properties. The objective is to review data mining models that can handle classification, prediction and association of high dimensional data with interactions, contain numerical responses, and that can model complex data structures. However, in addition to the previously mentioned objectives there is an underlying objective, one that is essential in any research, identify the areas of opportunity and research gap from previous models and methods in order to provide a novel contribution to the field.

#### **2.1 Data Mining**

Data mining consist of collection of techniques used for efficient automated discovery of previously mysterious, valid and useful patterns in large databases. It is increasingly used in the sciences to extract information from numerous data sets generated by experimental and observational study than any other fields of research (Hendricks; 2000). Data mining strategies can be categorized into: Classification, Estimation, Prediction, Association rule mining and Clustering (Hastie et al., 2001).

#### **2.2 Data Mining Models**

The various data mining models that are commonly applied to various problem domains include the Decision trees (Breiman et al., 1984), Artificial Neural Networks (McCulloch and Walter., 1943; Yahya et al, 2012), MARS (Friedman, 1991), Random Forest (Tim, 1995), Support vector machines (Cortes et al., 1995), Regression based classifier, K-SS and MK-SS classifiers (Yahya, 2009 & 2012; Yahya et al., 2011, 2014; 2015; etc.). Among these numerous methods, the

procedures of decision treat via CART, Artificial Neural Networks and MARS are briefly discussed here while their extension to modelling survival data are provided in Chapter Three.

### 2.2.1 Decision Trees

The most popular classification method is decision tree. The structure of the method is in form of a tree where internal node represents a decision on an attribute value. Each leave denotes an outcome of the decision and the leaves represent the classes. Decision tree is a model that is both analytical and descriptive. In data mining techniques, a decision tree is a prognostic model; that maps observations into their target value (Muehlenbrok, 2005). Additionally, expressive names for such tree models are classification tree (discrete response) or regression tree (numerical response). The discriminative power of each attribute that can best split the dataset is done either using the concept of information gain, Gini index or misclassification error. Popular decision tree algorithms like ID3, C4.5, C5 etc use information gain to select the next best attribute whereas popular packages like CART, IBM intelligent miner etc. use the Gini index concept (Burlak et al., 2006). The general algorithm for CART is as follows:

Let  $G_t$  denote the set of training information that extend to a node  $t$

- If  $G_t$  contains information belonging to class  $y_t$ , then  $t$  is a leaf node and represented by labeled as  $y_t$
- If  $G_t$  contains information that belong to two and more classes, then an attribute test is used to “split” the data into smaller subsets.
- The process is recursively applied to each subset.

The splitting criteria can be one of the following:

- i. *Misclassification error as splitting criteria*

$$Error(t) = 1 - \max P(i/t)$$

ii. *Gini index as splitting criteria*

$$GINI(t) = 1 - \sum_i [P(i/t)]^2$$

iii. *Information gain as splitting criteria*

$$Gain(t) = - \sum_i P(i/t) \log_2 P(i/t)$$

Figure 2.1 shows the partial view of a sample decision tree. One of the decision rule it provides is, if Outlook is sunny and humidity is normal, one can go ahead to play.

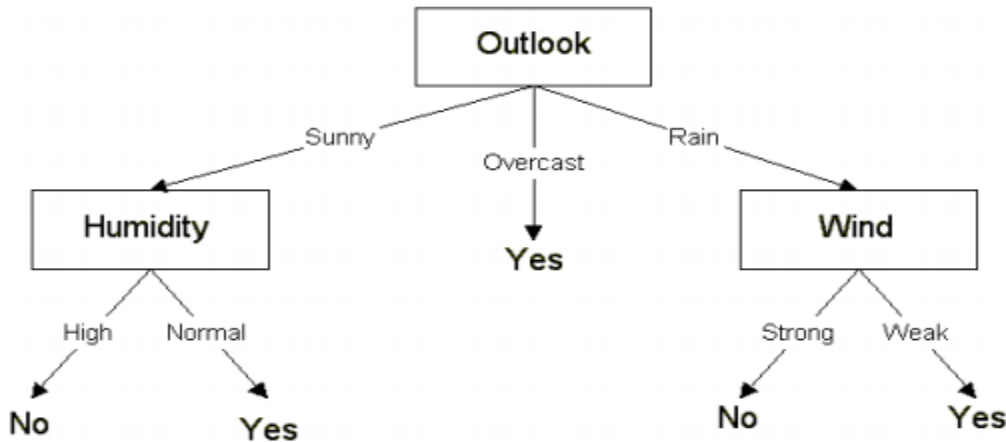


Fig. 2.1: Sample decision tree-Partial view

### 2.2.2 Neural Networks

A mathematical model that attempts to mimic the human brain is regarded as Neural networks (Waiyamai, 2003). Information represented in layers with a set of interconnectivity processors is known to be neurons in the anatomy of the human brain. At each leaf (node), there is a weight connected to other leaves (nodes) in head-to-head manners. Each node receives information as an input from connected nodes and the weights together with a simple function is used to compute the expected output values. Network connectivity is being used as learning algorithm in neural through the change in weight, and a set of inputs is repeatedly passed through the network. Once trained,

an unidentified instance passing through the network is categorized according to the values seen at the output layer (Hwang et al., 2008 and Zaïane et al., 2001).

Typically, a neural network model has a configuration as shown in figure 2.2 in its basic form. Neurons only fire when input is bigger than some threshold. It should, however, be noted that firing doesn't get bigger as the stimulus increases, it is an all or nothing arrangement (Atwell et al., 2006).

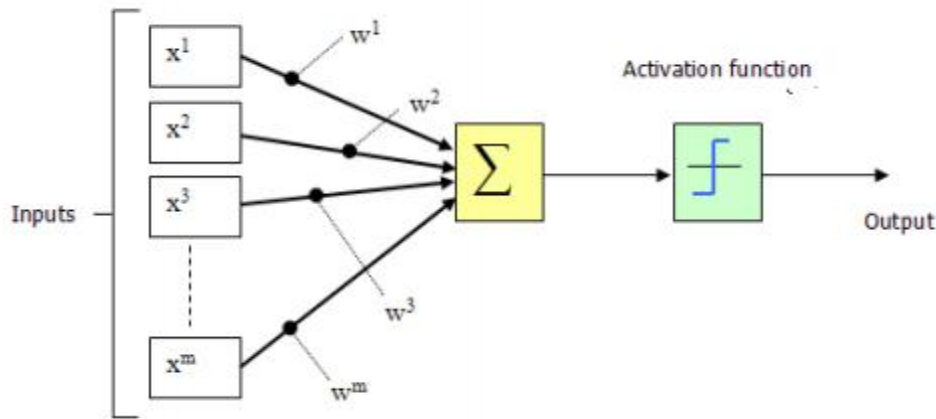


Fig. 2.2: A neural network structure

The structure shows that a neuron connects with mother neurons and so receives *m-many* inputs  $x^1, \dots, x^m$ . This configuration is actually called a Perceptron. A Perceptron models (Proposed by Rosenblatt, 1992) is a neuron by taking a weighted sum of inputs and sending the output, if the sum is greater than some adjustable threshold value otherwise it sends 0. This is the all or nothing spiking described in the previous paragraph. It is also called an activation function.

The inputs  $(x^1, x^2, x^3, \dots, x^m)$  and connection weights  $(w^1, w^2, w^3, \dots, w^m)$  in Figure 2.3 are typically real values, both positive (+) and negative (-). If the feature of some  $x^i$  tends to cause the perceptron to fire, the weight  $w^i$  will be positive; if the feature  $x^i$  inhibits the perceptron, the

weight  $w^i$  will be negative. The perceptron itself consists of weights, the summation processor, and an activation function, and an adjustable threshold processor (called bias).

For convenience, the normal practice is to treat the bias, as just another input. Figure 2.3 illustrates the revised configuration with bias. The bias can be thought of as the propensity (a tendency towards a particular way of behaving) of the perceptron to fire irrespective of its inputs.

The perceptron configuration network shown in Figure 2.3 fires if the weighted sum  $> 0$ , or in mathematical terms, it can be represented as in (2.1).

$$\text{Weighted sum} = \sum_i \text{bias} + (w^i x^i) \quad (2.1)$$

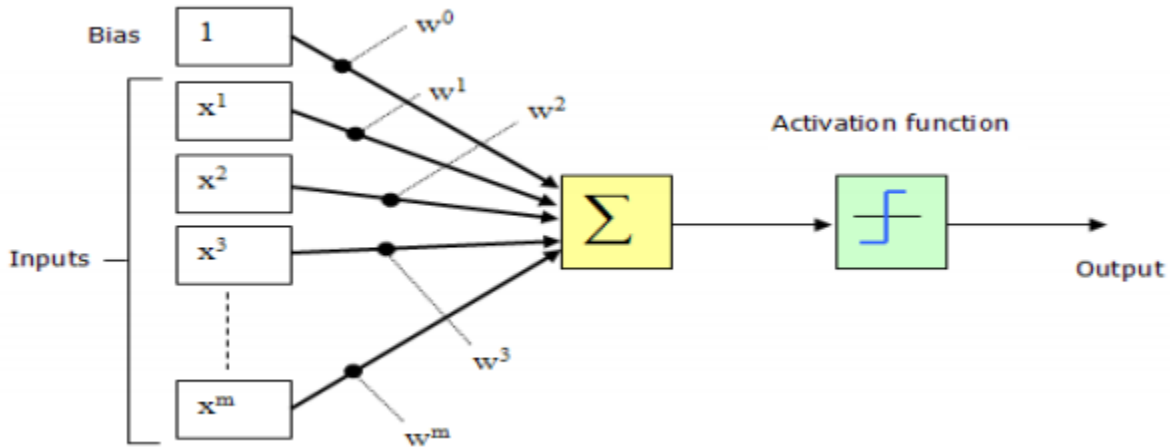


Fig. 2.3: Artificial Neuron configurations, with bias as additional Input

**Activation function:** An activation function performs a mathematical operation on the signal output. The most common activation functions are:

- i. **Sigmoid function:** The stronger the input is, the faster the neuron fires. The sigmoid is also very useful in multi-layer networks, as the sigmoid curve allows for differentiation (which is required in Back Propagation training of multi-layered networks). It is mathematically denoted as in (2.2)

$$f(x) = \frac{1}{(1+e^{-x})} \quad (2.2)$$

- ii. *Step function*: A step function is a basic on/off type function, if  $0 > x$  then 0, else if  $x \geq 0$  then 1. Hence depending on the type of input, output and problem domain, suited functions are adopted at respective layers.
- iii. *Piecewise Linear Function*: This activation function is also called saturated linear function and can have either binary or dipolar range for the saturation limits of the output.

The activation functions are chosen depending upon the type of problem to be solved by the network.

**Neural network architectures:** An artificial neural network (ANN) is a data processing system, consisting large number of simple highly interconnected processing elements as artificial neural in a network structure that can be represented using a directed graph. Among the several neural network architectures, for supervised classification, feed forward multilayer network trained with back propagation algorithm is the most popular. Neural networks, in which signal flows from input to output (forward direction) are called, feed forward neural networks. Other neural network architectures include: Recurrent Network and Single Layer Feed-forward Network (Bruce and Craig., 1999).

### 2.2.3 Multivariate Adaptive Regression Splines (MARS)

The MARS was developed by Friedman (1991) as statistical methods that own various properties that make it suitable for many applications, and for this reason, MARS has gained popularity over the last decades. It is a flexible method that is capable of fitting both linear and nonlinear functions without prior knowledge of the properties of such function. Many researchers such as: Kuhnert and McClure (2000), Psychogios et al. (1992) and Ben-Ari and Steinberg (2007) regard MARS as a nonparametric statistical method. MARS can model curvature by bending the function at certain knot locations, which are appropriately selected, and simultaneously choose the most significant

variables to be included in the final model. The approximation model is presented with the following equation:

$$\hat{f}_M(x; \beta) = \beta_0 + \sum_{m=1}^M \beta_m B_m(x) \quad (2.3)$$

where  $\beta_0$  is the intercept term,  $B_m(x)$  represents the basis functions that depend on the  $x$  factor variables, and  $\beta_m$  is the unknown coefficient for such functions.  $M$  represent the maximum number of basis function. The basis functions can be univariate or of a higher degree. The interaction terms are created by multiplying an existing basis function by a new univariate term. To represent the univariate terms, MARS uses piecewise-linear functions of the form:

$$b^+(x; k) = [+(x - k)]_+, \text{ or } b^-(x; k) = [-(x - k)]_+ \quad (2.4)$$

where  $k$  represents an univariate knot which is selected for each of the factor variables  $x$ . The functions are truncated at zero, and thus they only consider the positive side. These functions are known as truncated linear functions or hinge functions. The interaction functions are then represented by:

$$B_m(x) = \prod_{l=1}^{L_m} [s_{l,m}(x_{v(l,m)} - k_{l,m})]_+ \quad (2.5)$$

where the degree order  $L_m$  of the interaction term is user defined.

The execution of the MARS algorithm is as follows:

- i. **Stage-wise forward selection Algorithm:** Keep the coefficient the same for variable that existed in the current model, select a new basis function pair that produces the largest decrease in the training error. Repeat the whole process until some termination condition is met, for instance:
  - a. reached the maximum number (pre-determined) of predictors;
  - b. adding any new predictor changes the R-Square by less than 0.001;
  - c. Reached a very high R-Square value.

- ii. **Backward Algorithm:** For preventing over fitting, redundant basis functions are deleted from the model. MARS adopts Generalized Cross-Validation (GCV) to delete the redundant basis functions. The expression of GCV is given by:

$$GCV = \frac{\frac{1}{N} \sum_{i=1}^N [Y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(M)}{N}]^2} \quad (2.6)$$

where  $N$  is the number of data points and  $C(M)$  is the complexity penalty that increases with the number of basis function  $B_m(x)$  in the model and is defined as:  $C(M) = (B + 1) + dB$ , where  $d$  is a penalty for each  $B_m(x)$  included in the model and  $B$  is the number of the basis functions in the model. Hastie et al. (2001).

### 2.3 Related Research in Survival Tree-based with and without Competing Risks

Ibrahim and Kudus (2009) applied classification and regression tree into multivariate survival data based on marginal distribution. The generalization of decision tree for multivariate survival-tree was developed using goodness of split approach. In the process of growing tree, their developed approach was used to split the covariates by minimizing a measure of between-node differences in bladder tumor cancer data set. Their study presented not very high but considerable prediction accuracy rates. The main limitation of the study was: restriction to between-node differences, restriction to Wald statistic and utilization of a small dataset.

Ibrahim et al., (2008); used the classical approach aspect of the CART by utilizing the proportional hazards model for sub-distribution. They applied their technique to the breast cancer data where several events might occur following the treatment after breast-conserving surgery. Their study has two major limitations: First, they only considered splitting on a single covariate. Secondly, the

variables were selected based on the between-node experiences solely on survival analysis without competing risks.

Fiona (2008) proposed classification and regression tree for analyzing time to event data with competing risk outcomes by considering both univariate and multivariate techniques respectively. She developed within-node and between-node trees for competing risks data. Her proposed univariate tree for between-node tree was based on the difference between cumulative incidence functions, using Gray (1979) as measures of impurity and event-specific martingale residuals as measures of impurity in within-node tree. These methods addressed the problem of identifying risk of subgroups when there is more than one competing risk event. The major limitations of this study are: they utilized only one statistical technique (i.e. Gray) and ignored other important test statistics, their techniques were restricted to survival analysis with competing risks, and their major goal was to allocate organ to the patients on the waiting list while they ignored other prognostic factors that can be dangerous to the patients.

Yuhong et al. (2006) developed Bayesian computational method for Classification and Regression Tree (CART) model. Their modeling innovations include a formal prior distribution structure for tree generation known as “pinball prior” that allows the combination of an explicit specification of a distribution for both the tree size and the tree shape. Moreover, they used the new Metropolis-Hastings for their computation. They evaluated the effectiveness of the proposed algorithm in two examples, one with a constructed (simulation) data set and one concerning analysis of a published breast cancer data set. The major drawback of this study is their restriction to classical regression analysis, and non-survival data whether with or without competing events.

Chipman et al. (1998) proposed a stochastic search form of classification and regression tree motivated by a Bayesian model. They explored the use of an approximation to a probability distribution over the space of possible trees using reversible jump Markov Chain Monte Carlo Methods. A Bayesian model was set up for the distributions of the data points in the terminal nodes, using different forms for classification tree and regression tree problems. Furthermore, they generate samples from the joint posterior of  $(k, \theta^{(k)})$  by using a class of reversible jump Metropolis-Hastings algorithm. However, their drawback is like Yuhong et al. (2006), their proposed technique is only applied to classical regression analysis, without considering survival analysis scenario.

Ishwaran et al. (2014), introduces survival random forest into survival analysis with competing risk events. This technique was fully non-parametric for high dimensional survival data on HIV/AIDS that contains many competing risks events. The method was used to select event specific variables and for estimating the cumulative incidence function. The major limitations of this study are that their method was purely based on non-parametric approach and they only provided probability of survival via cumulative incidence function.

Kretowska (2012) proposed a dipolar ensemble tree for analyzing competing risks. This method was built under the learning sets. The proposed technique has the capability of accounting for incomplete censoring in the data set via the cumulative incidence function. The limitation of this technique is similar to Ishwaran et al. (2014), where only probability of survival was estimated via their technique and small data set was also utilized.

Orna and Kooperberg (1995) developed a method based on trees and splines (polynomial), to analyse two data using Classification and Regression Tree in survival analysis and Cox

Proportional Hazard (CHS). The strength and weaknesses of the two methods were compared and tested using model dimension, BIC-value and Log-likelihood based on the Coronary heart disease data. The two major limitations of their study are: Firstly, they consider only survival analysis without competing risks, and log-rank was used as a measure of separation between the nodes, other important statistics were ignored. Secondly, their methods only considered a basis function with polynomial of order two, instead of allowing the techniques to select the best splines.

Wei and Jeffrey (2016), proposed Classification and Regression Tree in survival analysis for left and right truncated which can be used as traditional survival tree-based method. The methods were further extending to time varying covariates. Both simulation and real-life data were used to validate the performance of the proposed methods. The major drawback of this study is: they only applied their method to survival analysis without competing risks and only between-node separation was considered.

Feng et al. (2006), a tree-structural method was developed to extend the Classification Regression Tree (CART) algorithm to multivariate survival data. The proportional hazard structure was assumed to take effect on the whole tree. The method works directly on the marginal survivor distribution that uses a sandwich estimator of variance and accounting for association between survival times. The proposed method was applied to a prostate cancer and evaluated its performance by several simulations studies. The only limitation of this study is that only the Wald test statistic was used as splitting rule and the survival tree was developed by maximizing between-node separation.

## **2.4 Related Research in Multivariate Adaptive Regression Spline Survival Analysis Based**

In 2007, Kriner proposed a new model for modelling survival time data with Multivariate adaptive regression spline (Kriner, 2007). The martingale and deviance residuals of a Cox PH model were used as responses in a common MARS approach to model functional forms of covariate effects as well as possible interactions in a data-driven way. She performed different simulation studies to the newly developed methods and compares her results with the existing classical method, and the new methods yield a better fit to the data than the existing method. She further applied a real data of the German Heart Center on survivors of an acute myocardial infraction and the results revealed a good performance of her proposed methods over the classical methods. However, the major limitation of this study is that the deviance and the martingale she considered can only be used mainly for a data set that is heavily skewed and failed when skewed to both sides.

Kooperberg et al. (1997), proposed an automatic procedure that uses linear splines and tensor product for fitting classical regression model to data consisting a polychotomous response variable and one or more predictors. This procedure consists of using maximum likelihood, stepwise addition, stepwise selection and model selection by using some performance indices such as Akaike Information Criteria (AIC) and cross-validation procedure. Unfortunately, this technique is only applied to classical regression model and polychotomous response, but not in survival analysis. This formulates the major drawback of their technique.

LeBlanc and Crowley (1999), developed a method for modelling adaptive regression spline in survival analysis. The method utilizes both Cox regression model and weighted least square for multivariate adaptive regression spline, to adaptively select knot and covariate in the data set. The proposed model was applied to simulation data and clinical trial for myeloma data set. The major

limitation of this study is that: their methods is restricted to likelihood estimate of the CPH, and focus majorly on the variable selection based on adaptive spline. The interaction effect of the covariates is not considered but rather ignored during their simulation processes.

Costa and Shaw (2006), built a cubic spline that aids the implementation and interpretation of penalized spline, irrespective of the selected knot. This approach was called value-first derivative parametrization. Bayesian inference by exploring the natural link between quadratic penalties and Gaussian priors was adopted and an empirical Bayes method was used as a model selection type criteria. The proposed model application was performed on survival real-life data. Their techniques are strictly on Bayesian inference and no consideration is given to the model when prior is not available, particularly on a real-life data. Additionally, their model is built majorly on classical regression analysis, no consideration on survival data. All these points formulate the major drawback of their study.

Bani et al. (2000), adopt the use of Bayesian multivariate adaptive regression spline approach to model univariate and multivariate survival analysis data with censoring. The model contains proportional hazard model as a subclass, and automatically detect departure. The algorithm used to estimate the hazard function as well as survival curve was based on reverse Markov chain Monte Carlo. The major drawback of this techniques is similar to Bani et al. (2000), where the methods are strictly on Bayesian approach on classical regression analysis only.

## **2.5 Related Research on Artificial Neural Network in Survival Analysis**

Chih-Lin et al. (2007), adopt the theory of Artificial Neural Networks (ANNs) to model time to event data problems. Three layers, feed-forward ANNs with sigmoid activations function was applied and the network was trained using standard back-propagation algorithm which contains

three layers, an input layer, a hidden layer, and an output layer. The results of the model were compared with Kaplan Meier estimate based on two cancer data set. The main limitation of the study was the utilization of survival data without competing risk and no major statistical measure of indices was used to compare their results.

Malgorzata and Leon (2000), designed artificial neural network for prediction of survival time. The main aim of the method was to identify areas in the feature space of homogenous in the survival analysis perspective. Moreover, the method is based on minimization of a piece wise linear function and dipolar criterion function designed to account for censoring in the data set. This study has two major limitations: First, they lack an enhanced data-mining perspective which would utilize machine learning and artificial intelligence tools to reveal the previously-unknown potentially-useful patterns. Secondly, the variables were selected based on dipolar criterion, which cannot account for the interaction effects among the covariates.

Also, Jonas et al. (2013) developed a prognostic index model for survival analysis based on an ensemble of artificial neural networks that optimize directly from the concordance index. Approximations of the c-index are avoided with the use of a genetic algorithm, which does not require gradient information. The model is compared with Cox proportional hazards based on some performance indices. This study is restricted to survival data analysis without competing risks event and no consideration is given to a with-competing-risk scenario, and this forms the major limitation of this study. A more lately held study has the same drawbacks, which was proposed by Yanying (2010).

Yanying (2010), applied the concept of artificial neural network to a survival data set. The data set was analyzed by a piecewise standard Cox PH model, a piecewise marginal Cox model and the

PLANN neural network approach. The performance of these methods were compared using hazard ratio and survival probabilities. However, similar to the aforementioned studies, their limitations are the same but with addition of using the marginal CPH model.

Kalderstam (2015) explored different methods to predict the patient's survival chances using artificial neural networks (ANN). The techniques utilize a genetic algorithm to train ANN models to maximize (or minimize) non-differentiable functions, which are impossible to combine with traditional ANN training techniques which rely on gradient information. The methods does not use a genetic algorithm approach, rather they take the approach to modifying the underlying data. Regular gradient methods used to train ANNs on survival data were estimated in a maximum likelihood framework. The major limitation of this study is that: their study is restricted to survival data analysis without competing events, and no consideration is given to competing events. Their techniques also applied to a small survival dataset with just one events of interest.

## **2.5 Summary and Research Gaps**

The classical methods for analyzing survival data without competing risks include: Parametric Models (i.e. modelling survival data with exponential, Weibull and Lognormal distributions), Cox Proportional Hazard (CPH), and Kaplan Meier. While survival data with competing risks can be modelled with the following methods: Cumulative Incidence Function and Cause-Specific Hazard. However, the main limitations of the aforementioned studies are that they do not give satisfactory results at the local or regional levels when the underlying assumptions that govern the usability of these models fail when the survival data is of high dimension (i.e. the number of observation is less than the number covariates). These refer to a major gap in modelling of survival patient with

Cancer disease (or other diseases). Therefore, if a well-established decision support system that would determine good indicators of medical urgency/priority through data mining-based survival and prognostic analyses can be developed; there will be a linkage to better simulation scenarios and results at all levels of disease prognostic risk factors. Similarly, such a comprehensive methodology could help incorporate more survival outcome of the disease at various levels of risks. This research study is intended to cover all these aforementioned gaps and pave way to overcome these challenges as further explained in Chapter Three.

## CHAPTER THREE

### METHODOLOGY

This work adapts integrated data mining methods to model the complex relationship between predictor variables and survivability. The techniques to be extended to survival based-tree are: Classification Regression Tree (CART) via within-node homogeneity at the first step (Task 1) and between-node heterogeneity at the second step (Task 2), Multi-layer Perceptron Artificial Neural Network (MLPANN) at the third step (Task 3) and Multivariate Adaptive Regression Spline (MARS) at the last step (Task 4). A pictorial representation of the overall Classification Regression Tree (CART) methodology is illustrated in Figure 3.1. These four tasks are further explained in the following sections.

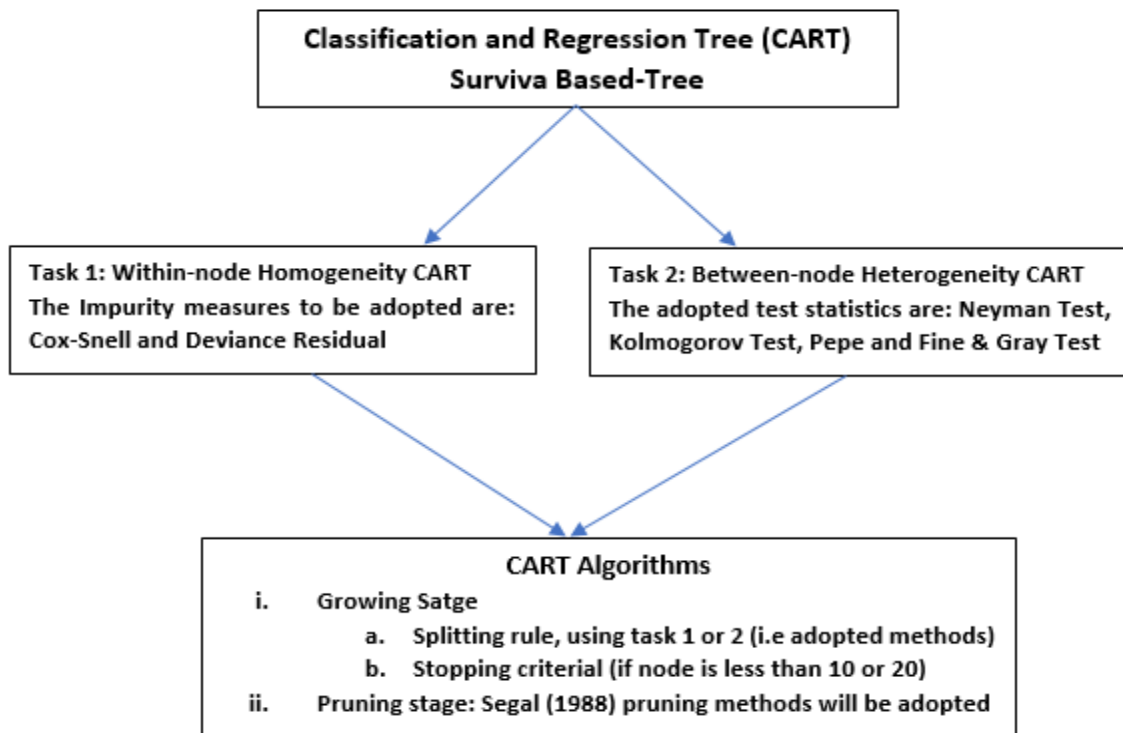


Fig. 3.1: Framework for Classification and Regression Tree (CART) research methodology

### 3.1 Classification and Regression Tree (CART) Methodology

A simplified description of CART is presented in this subsection, so that the subsequent reformulations can be understood. Attention here is centered on the familiar regression setting (Linear Regression Analysis). The Classification and Regression (CART) technique of Breiman et al. (1984) addresses the recursive partitioning issue by constructing a binary decision tree according to some splitting rule based on the covariates. In this way, the space of predictor variables is partitioned recursively in a binary way. The partitioning is intended to maximize within-node homogeneity (or between-node heterogeneity), where homogeneity (or heterogeneity) is determined by dependent variables in the problem. The partitioning is repeated until a node is reached for which no split improves the homogeneity (or heterogeneity), whereupon the splitting is stopped and this node becomes a terminal node. Prediction is determined by terminal node, and is either a class label in classification problems or the average of the response variable in Least Squares (LS) or Least Absolute Deviations regression problems. Two stages are involved in constructing a tree: Growing and Pruning. However, the methods of constructing a regression-tree (i.e. growing stage) constitute four components, which are:

1. Partitioning or split on the predictor space (greedy algorithm)
2. A goodness of split criterion  $\phi(s, t)$  that can be evaluated for any split  $s$  of any node  $t$ . The criterion is used to assess the worth of the competing splits pertaining to within-node homogeneity (or between-node heterogeneity).
3. A mean of determining appropriate tree size
4. Summary of the terminal node of the selected tree.

What follows is an elaborate of this aspect (Growing stage)

### 3.1.1 General Algorithm to Grow Tree

The impurity functions for every possible binary split of the predictor space  $Z$  will be evaluated. The split  $s$  could be of several forms: splits on a single covariate, splits on linear combinations of predictors, and Boolean combination of splits. The simplest form, in which each split relates to only one covariate, can be described as follows:

1. For continuous/ordinal covariate  $Z$ , all the possible cut points'  $c$  that split the observation into two distinct categories is determined. i.e.  $Z \leq c$  and  $Z > c$ .
2. And for a nominal covariate, all the possible ways of forming two categories were considered.

These processes will be repeated for all subsequent branches of the tree until the best split are achieved using the stopping criteria. The traditional stopping criteria are that the sample size in each terminal node is very small and or the subgroups are homogeneous with respect to the covariates. Following these procedures recursively together with the best split above, the appropriate large tree size is determined and could be used for the purpose of data structure (i.e. Tree diagram). For point of reference and specificity the definition of the goodness-of-split on the Least Squares (criterion) is presented below according to LeBlanc and Crowley (1993).

Let  $t$  denote a node of the tree. That is,  $t$  contains a subsample  $\{(\vec{x}, y_n)\}$ . Let  $N(t)$  be the total number of cases in  $t$  and let

$$\bar{y}(t) = \frac{1}{N(t)} \sum_{\vec{x} \in t} y_n \quad (3.1)$$

be the node response average. Then the within-node sum-of-squares is given by equation (3.2) below

$$SS(t) = \sum_{\vec{x} \in t} (y_n - \bar{y}(t))^2 \quad (3.2)$$

Now suppose a split  $s$  partition  $t$  into left and right daughter nodes  $t_L$  and  $t_R$ . Therefore, LS criterion is given in equation (3.3)

$$\phi(s, t) = SS(t) - SS(t_L) - SS(t_R) \quad (3.3)$$

and the best split  $s^*$  is the split such that

$$\phi(s^*, t) = \max \phi(s, t) \quad (3.4)$$

where the risk at the node  $t$  is given by  $SS(t) = p(t)s^2(t)$  where  $p(t)$  is the proportion of observations that fall in  $t$  and  $s^2(t)$  the variance of the corresponding values of the dependent variables. A useful feature of CART is that after growing a large tree, a pruning will be done to get a sequence of nested pruned subtrees. This is done using a penalty measure based on the complexity of the tree and called cost complexity. Thus, the general algorithm for pruning tree is given in the next subsection.

### 3.1.2 Algorithm to Prune Tree

The idea of pruning is to iteratively cut off branches of the initial tree, in order to allocate a limited number of candidate sub-trees from which an optimally sized tree is selected. For the purpose of this study method, the Segal's pruning algorithm (Segal, 1988) which exerts little computational burden would be considered. The steps for adopting this algorithm are as follows:

- i. Initially growing a large tree.

- ii. To each of the internal nodes in the original tree, assign the maximal splitting statistics contained in the corresponding branch. This statistic reflects strength of linking from the branch to the tree.
- iii. Among all these internal nodes, find the one with the smallest statistic. That is, find the branch that has the weakest link and then prune off this branch from the tree.
- iv. The second pruned tree can be obtained in a similar manner by applying the above two steps applied to the first pruned tree.
- v. Repeat the process until the pruned tree contains only the terminal node, finally a sequence of nested trees is obtained.

In subsection 3.1.3 and 3.1.4, the extension of CART into survival analysis (survival tree-based) via within- and between-node approaches will be discussed. The extension is quite different from the traditional regression briefly explained in the previous section and subsections since survival data contain both time and status as a response. The sum-of-squares and absolute deviation explained earlier will be of help and use for a different purpose. The main extension of this study focuses on the splitting criteria which is the back bone of CART, the development of such criteria will be explained in detail under the growing stage of tree. However, other procedures of CART algorithm remain the same throughout this study.

### **3.1.3 Task 1: Survival Tree-Based via within-node homogeneity**

In survival analysis with competing risks, under the univariate event, a tree will be built by concentrating on the event of interest and adjusting for other competing risks. The study proposed to use within-node homogeneity by using Deviance, Cox-Snell and Martingale residuals in sub-distribution hazard function developed by Fine and Gray (1999) i.e. hazard function associated

with Cumulative Incidence Function (CIF). The sub-distribution hazard function  $\tilde{\lambda}(t)$  is assumed to take the form:

$$\tilde{\lambda}(t; Z_i) = \tilde{\lambda}_{j0}(t) \exp[\beta^j I(Z_i < c)]$$

Where  $\tilde{\lambda}_{j0}(t)$  is an unspecified baseline sub-distribution and  $\beta^j$  is an unknown parameter corresponding to cut-point  $c$  (Ibrahim et al., 2008). The splitting rule (i.e. impurity function) via Deviance, Cox-Snell and Martingale (used by Fiona, 2008) residuals will be rooted into the general algorithm of CART. The general algorithm of the CART has been fully discussed in section 3.1.3.1.

#### ***3.1.3.1 Growing of Survival Tree-Based via within-node homogeneity***

The required homogeneity measure that is appropriate to design a survival tree-based data with competing risks are employed, and they are: Cox-Snell, Deviance and Martingale residuals. But the Martingale residual has been used by Fiona (2008). Therefore, Cox-Snell and Deviance residuals will be modified for use in splitting procedure for this study. However, the performance of the proposed methods (Cox-Snell and Deviance residuals) will later be compared with the existing method (Martingale residual) through the cutpoint detection and tree structure. The modification of these homogeneity measures (i.e. Cox-Snell and Deviance residuals) will be in the following deviations: Sum-of-squares and absolute value impurity functions. This impurity functions will be taken one after the other for further explanation.

##### **A. Sum-of-squares impurity function**

This study proposes the use of Cox-Snell (CS) and Deviance (D) residuals as measures of impurity in  $m$  competing risk. For individual  $i$  in group  $k$  for any event  $m$  is defined as:

i. Using Cox- Snell Residual (CSR)

$$\widehat{CSR}_{ik}^m = I_{\{\delta_{ik}=m\}} - \widehat{\Lambda}_0^m(T_{ik}) \quad (3.5)$$

where  $m$  is the outcome measure on the event ( $m = 1, 2, 3, \dots, J_{\text{risk}}$ ) and  $\widehat{\Lambda}_0^m(\cdot)$  denote the value of estimated cause-specific hazard function, and the function (cause-specific hazard function) can be calculated as follows by equation (3.6), and equation (3.5) is the estimate of the Cox-Snell residual

$$\widehat{\Lambda}_0^m(T_{ik}) = \int_0^{T_{ik}} \frac{dN_{mk}(s)}{Y_k(s)} \quad (3.6)$$

where  $N_{mk}(t)$  and  $Y_k(t)$  denote the corresponding number of event  $m$  and possible risks at time  $t$  for group  $k$ , respectively.

For a node  $h$ , the study proposes using an impurity function  $\phi_{ss}(h)$  based on the sum of the square, for Cox-Snell residual is given by equation (3.7).

$$\phi_{CSRSS}(h) = \sum_{i \in h} (\widehat{CSR}_{ih}^m - \overline{CSR}_h^m)^2, \quad (3.7)$$

Let  $s$  be a possible split for the parent node  $P$ , and let  $L$  and  $R$  be the corresponding left daughter node and right daughter node, respectively, for this split. To obtain the split that maximizes the within-node homogeneity, the study needs to find an  $s$  (i.e. possible split) that has the largest reduction in within-node impurity from the parent node  $P$ . That is, there is need to find an  $s$  that maximizes the impurity function  $\Phi_{CSRSS}(s, P) = \phi_{ss}(P) - \phi_{ss}(L) - \phi_{ss}(R)$ . Further simplification was done on the process, since  $\phi_{CSRSS}(s, t) = p(t)s^2(t)$  where  $p(t)$  is the proportion of observations that fall in  $t$  and  $s^2(t)$  is the sum-of-squares of the corresponding values of the dependent variables and the results yielded equation (3.8).

$$\Phi_{CSRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{CSR}_L^m - \overline{CSR}_R^m)^2 \quad (3.8)$$

This process is repeated for all subsequent nodes until we research the stopping criteria described in subsection 3.1.0. This recursive procedure grows the largest possible tree based on first proposed splitting criterion given by equation (3.8).

ii. Using Deviance Residual (DR)

The above steps and procedures were followed appropriately in other to produce the impurity function via the deviance residual using the sum-of-squares. The process is done accordingly, and the second splitting criterion is achieved and the result is given in equation (3.9).

$$\Phi_{DRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{DR}_L^m - \overline{DR}_R^m)^2 \quad (3.9)$$

iii. Using Martingale Residual (MR)

Similarly, the impurity function criteria of growing tree proposed by Fiona (2008) via martingale residual is given by equation (3.10).

$$\Phi_{MRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{MR}_L^m - \overline{MR}_R^m)^2 \quad (3.10)$$

The summary of the proposed and the existing splitting criteria based on sum-of-squares impurity functions is given in table 3.1 for better appreciation and justification.

Table 3. 1: Splitting rule criteria on sum-of-squares impurity functions

| S/N | Impurity functions                                                                     | Status       |
|-----|----------------------------------------------------------------------------------------|--------------|
| 1   | $\Phi_{DRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{DR}_L^m - \overline{DR}_R^m)^2$    | Proposed (1) |
| 2   | $\Phi_{CSRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{CSR}_L^m - \overline{CSR}_R^m)^2$ | Proposed (2) |

|   |                                                                                     |       |
|---|-------------------------------------------------------------------------------------|-------|
| 3 | $\Phi_{MRSS}(s, P) = \frac{N_L N_R}{N_P} (\overline{MR}_L^m - \overline{MR}_R^m)^2$ | Exist |
|---|-------------------------------------------------------------------------------------|-------|

### B. Absolute value impurity function

The most commonly used methods in the literature is the sum-of-squares or  $L2$  metric. However, survival data or any skewed data is often better summarized using an absolute value because is less susceptible to extreme values. Therefore, this study also proposes using the following impurity functions summarized in Table 3.2 based on the absolute value function. The methods were developed following the procedure like the one under sum-of-squares by taking the impurity function for Cox-Snell residual as example to be  $\phi_{CSRSS}(s, t) = p(t)|s(t)|$  respectively.

Table 3. 2: Splitting rule criteria on Absolute value impurity functions

| S/N | Impurity functions                                                                     | Status       |
|-----|----------------------------------------------------------------------------------------|--------------|
| 1   | $\Phi_{DRAV}(s, P) = \frac{N_L N_R}{N_P}  \widehat{DR}_{ih}^m - \overline{DR}_h^m $    | Proposed (3) |
| 2   | $\Phi_{CSRAV}(s, P) = \frac{N_L N_R}{N_P}  \widehat{CSR}_{ih}^m - \overline{CSR}_h^m $ | Proposed (4) |
| 3   | $\Phi_{MRAV}(s, P) = \frac{N_L N_R}{N_P}  \widehat{MR}_{ih}^m - \overline{MR}_h^m $    | Exist        |

#### 3.1.3.2 The Pruning of Survival Tree-Based via within-node homogeneity

In order to adjust for over fitting of the impurity function  $\phi(h)$  described above, tree grown has to be pruned. We used the cost complexity algorithm presented by Breiman et al. (1984) to obtain an ordered list of optimally pruned sub-trees, substituting our impurity function for node h,  $\phi(h)$ , based on event-specific residuals.

The cost-complexity function of tree T is given as:

$$\phi_\alpha(T) = \sum_{h \in T^t} \phi(h) + \alpha |T^t| \quad (3.11)$$

where  $\alpha$  is a non-negative complexity parameter.

The tree that minimizes the cost-complexity function (3.11) is the optimally pruned sub-tree. As  $\alpha$  increases, there are “jumps” whereby a smaller sub-tree becomes the minimizing tree for a new value of  $\alpha$ , say  $\alpha^*$ . The  $\alpha^*$ ’s is defined as

$$\alpha^* = \frac{\phi(h) - \phi(T_h)}{|T_h^t| - 1} \quad (3.12)$$

where  $\phi(T) = \sum_{h \in T^t} \phi(h)$  and  $T_h$  represent the root node  $h$  of the tree. The subdivision with the smallest cost complexity is pruned recursively until we have a decreasing sequence of smallest optimally pruned sub-trees.

$$T_m < T_{m-1} < \dots < T_1 < T_0$$

and the arrangement of corresponding  $\alpha^*$ ’s is

$$\infty = \alpha_m^* > \alpha_{m-1}^* > \dots > \alpha_1^* > \alpha_0^* = 0$$

In other to select the final tree, cross validation method was adopted and grows an optimally pruned tree through the sequence of trees  $T_m < T_{m-1} < \dots < T_1 < T_0$  on the entire data  $\mathcal{L}$ . Then, the data set is divided as  $\mathcal{L}_{(v)} = \mathcal{L} - \mathcal{L}_v$ . For each  $v$ , we use  $\mathcal{L}_{(v)}$  to grow a nested sequence of optimally pruned trees.

$$T_m^v < T_{m-1}^v < \dots < T_1^v < T_0^v, \text{ which we obtain by pruning } T_0^v \text{ using } \alpha'_m = \sqrt{(\alpha_m^* \alpha_{m+1}^*)}.$$

In place of individual tree  $T_m^v$ , the sum of the squared Cox- Snell and Deviance residual for the data set  $\mathcal{L}_v$  is calculated, with aim of getting more accurate estimate of the overall impurity of

individual grown tree. By averaging, summing over  $V$  sets and selecting the tree that minimizes the impurity, then adjusting for the number of terminal nodes. The total impurity function given in equation (3.12) is known in advanced. Then, by letting

$\phi(\mathcal{L}_v; \mathcal{L}_{(v)}, T) = \sum_{h \in T^t} \sum_{i \in h} \{(\widehat{CSR}_i^m - \overline{CSR}_h^m)^2\}$  represent the results of  $\phi$  for tree  $T$  after  $T$  is grown with  $\mathcal{L}_2$  and then assessed on  $\mathcal{L}_1$ . We then can calculate

$\phi(\mathcal{L}_v; \mathcal{L}_{(v)}, T_m^v) = \sum_{h \in T^{vt(\alpha)}} \sum_{i \in h} \{(\widehat{CSR}_i^m - \overline{CSR}_h^m)^2\}$  for every value of  $v$  and take the average,

$$\phi^{CV}(T_m) = \frac{1}{V} \sum_{v=1}^V \phi(\mathcal{L}_v; \mathcal{L}_{(v)}, T_m^v).$$

The final tree,  $T(\alpha_c)$ , is selected from the real list of optimally pruned sub-division trees grown on the whole section  $\mathcal{L}$ . The  $T(\alpha_c)$  is the smallest optimally pruned sub-tree that has the minimum cross-validation adjusted  $\phi_\alpha(T)$  for  $\alpha = \alpha_c$ . This means it is

$$T(\alpha_c) = \arg_{T_m \in \{T_0, \dots, T_m\}} \phi^{CV}(T_m) + \alpha_c |T_m^t|$$

We apply this process to Deviance Residual. This process is similar to the one adopted by Fiona (2008) and LeBlanc et al. (1993).

### 3.1.3.3 Simulation Studies for Survival Tree-Based via within-node homogeneity

In this section, the full details of simulation scheme were discussed in order to investigate how accurately the adopted techniques detect cut-points on a covariate and how well the tree procedures detect the data structure. We compared the performance of proposed within-node homogeneity tree based on an event specific Martingale, Deviance and Cox-Snell residuals impurity function with sums-of-squares and absolute deviations presented in Table 3.1 and 3.2.

The survival times were simulated based on proportional distribution (exponential) and non-proportional distribution (log-normal).

#### i. **Cut-Point Detection Simulation Scheme**

The essence of the simulation is to determine and investigate the accuracy of the proposed methods to detect the correct cutpoint on the covariate and also check the performance in respect to data structure. Additionally, we compared our methods with the methods used by Fiona (2008). Two major distributions are of interest and these are: exponential and log-normal distributions, and the survival time were simulated based on these two distributions.

Tables 3.3 and 3.4 showed the setups for differentiating the models used for the simulation in which the Cumulative Incidence Function (CIF) for the two events (i.e. the competing risks) were dependent on a covariate  $Z$  at a known cutpoint  $C$ . The cutpoint was set to range between 0.1 to 0.9 and  $Z$  was set to follow a uniform distribution with minimum 0 and maximum 1. Each model was generated with different failure rate such as letting  $Z \leq C_i$  and  $Z > C_i$ , for  $i = 1$  or  $2$ . These rates were selected to approximate the real-life situation of patients with the disease (e.g. cancer disease). The censoring rates was categorized into three classes: Low censoring (25% i.e. first quartile), Moderate censoring rates (50% second quartile) and High censoring rates (75% Upper quartile) depending on the exponential and log normal distributions. For the log normal, the shape parameter  $\sigma = 1$  was used for every censoring rate.

The sample size for each observation is 200, and 2000 data set for each model were generated using Bootstrap technique. To compare the performance of proposed splitting rule, we used the following performance measures: Mean, Median, Standard Deviation (SD), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

$$Mean(\hat{c}) = \frac{\sum_{i=1}^{2000} \hat{c}_i}{2000},$$

$$RMAE = \sqrt{\frac{\sum_{i=1}^{2000} (c - \hat{c}_i)^2}{2000}},$$

$$MAE = \frac{\sum_{i=1}^{2000} |c - \hat{c}_i|}{2000}$$

Table 3. 3: Model for Data Structure Detection via Exponential Distribution

| Exponential distribution |                                             |                |                                             |                |
|--------------------------|---------------------------------------------|----------------|---------------------------------------------|----------------|
|                          | Cumulative Incidence Function (CIF) Event 1 |                | Cumulative Incidence Function (CIF) Event 2 |                |
|                          | Low censoring rate (25%)                    |                |                                             |                |
| Cut-point                | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$ | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$ |
| 0.1                      | Exp(0.2)                                    | Exp(0.4)       | Exp(0.7)                                    | Exp(0.5)       |
| 0.2                      | Exp(0.2)                                    | Exp(0.4)       | Exp(0.7)                                    | Exp(0.5)       |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| 0.9                      | Exp(0.2)                                    | Exp(0.4)       | Exp(0.7)                                    | Exp(0.5)       |
|                          | Moderate censoring rate (50%)               |                |                                             |                |
| 0.1                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.4)                                    | Exp(0.2)       |
| 0.2                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.4)                                    | Exp(0.2)       |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| 0.9                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.4)                                    | Exp(0.2)       |
|                          | High censoring rate (75%)                   |                |                                             |                |
| 0.1                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.7)                                    | Exp(0.5)       |
| 0.2                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.7)                                    | Exp(0.5)       |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| .                        | .                                           | .              | .                                           | .              |
| 0.9                      | Exp(0.1)                                    | Exp(0.3)       | Exp(0.7)                                    | Exp(0.5)       |

Table 3. 4: Model for Data Structure Detection via Lognormal Distribution

| Log-normal distribution |                                             |                  |                                             |                  |
|-------------------------|---------------------------------------------|------------------|---------------------------------------------|------------------|
|                         | Cumulative Incidence Function (CIF) Event 1 |                  | Cumulative Incidence Function (CIF) Event 2 |                  |
|                         | Low censoring rate (25%)                    |                  |                                             |                  |
| Cut-point               | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$   | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$   |
| 0.1                     | Log-normal(1.25)                            | Log-normal(0.5)  | Log-normal(0.1)                             | Log-normal(0.25) |
| 0.2                     | Log-normal(1.25)                            | Log-normal(0.5)  | Log-normal(0.1)                             | Log-normal(0.25) |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| 0.9                     | Log-normal(1.25)                            | Log-normal(0.5)  | Log-normal(0.1)                             | Log-normal(0.25) |
|                         | Moderate censoring rate (50%)               |                  |                                             |                  |
| 0.1                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.25)                            | Log-normal(1.25) |
| 0.2                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.25)                            | Log-normal(1.25) |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| 0.9                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.25)                            | Log-normal(1.25) |
|                         | High censoring rate (75%)                   |                  |                                             |                  |
| 0.1                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.1)                             | Log-normal(0.25) |
| 0.2                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.1)                             | Log-normal(0.25) |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| .                       | .                                           | .                | .                                           | .                |
| 0.9                     | Log-normal(2)                               | Log-normal(0.75) | Log-normal(0.1)                             | Log-normal(0.25) |

## ii. Data Structure Performance

We examine the ability of the adopted survival tree methods via the data structure on survival data with competing risks. In the Table 3.5, models 1 and 2 were generated from exponential distributions, and models 3 and 4 were also generated from log normal distributions. We let two predictors  $Z_1$  and  $Z_2$  to be associated to survival times, and four others  $Z_3, Z_4, Z_5$  and  $Z_6$  be independent of survival times. Covariates  $Z_1, Z_3$  and  $Z_4$  were made to be binary in nature and

simulated from binomial distribution with parameter  $p = 0.5$ . Whereas other covariates  $Z_2, Z_5$  and  $Z_6$  were simulated through the uniform distribution with the following ranges (parameters):  $[0, 0.25]$ ,  $[0.5, 0.75]$  and  $[0,1]$  respectively. This study considers 50% censoring rate for all the simulation on both additive and interaction models. Model 1 and 3 had an interaction (multiplicative) effect on  $Z_1$  and  $Z_2$ , so that a tree with three terminal nodes would be expected to be best data structure. Similarly, model 2 and 4 had additive hazards for the competing event  $i$  based on  $Z_1$  and  $Z_2$ , a tree with four terminal nodes would be expected to be the best data structure.

This study constructed 100 trees with 500 sample sizes for each simulation. To measure the performance of the proposed splitting rule, both the number of terminal nodes in each tree and the forecasting capability of the tree were considered. The generated 500 observations were used to validate the data sets without censoring for each simulation and estimate the mean absolute difference between the risk 1 failure time for each observation in a validation data set and the median of risk 1 failure time predicted by the tree. Similarly, this is also repeated to the second event 2 and below is the computation procedure for risk 1 and risk 2 times respectively, as

$$MAE_1 = \frac{1}{N_1} \sum_{h=1}^{|T^t|} \sum_{i=1}^{N_{1h}} |T_{i1h} - \hat{t}_{1h}|, \quad MAE_2 = \frac{1}{N_2} \sum_{h=1}^{|T^t|} \sum_{i=1}^{N_{2h}} |T_{i2h} - \hat{t}_{2h}|$$

Where  $N_{1h}$  represent the number of event 1 in node  $h$ ,  $N_1$  total number of event 1,  $T_{i1h}$  is the  $i$ th failure time for event 1 in terminal node  $h$ , and  $\hat{t}_{1h}$  is the median event 1 failure time based on the training data for terminal node  $h$ . This interpretation is also applicable to event 2. 1 is choosing as complexity penalty function, 20 numbers of minimum observations in each terminal node is allowed and 10-fold Cross Validation technique was used to select the final tree.

Table 3. 5: Description of models used for simulation in data structure performance

| Model | Cumulative incidence function<br>$F_1(t/Z_1, Z_2)$                           | Expected terminal<br>nodes<br>$ T^t $ |
|-------|------------------------------------------------------------------------------|---------------------------------------|
| 1     | <b>Exponential</b> $\lambda = 0.1 + 0.3I\{(Z_1 > 0.5) \cap (Z_2 > 0.5)\}$    | 3                                     |
| 2     | <b>Exponential</b> $\lambda = 0.05 + 0.2I(Z_1 > 0.5) + (Z_2 > 0.5)$          | 4 or 2                                |
| 3     | <b>Log normal</b> $\mu = 2 + 1.5I\{(Z_1 > 0.5) \cap (Z_2 > 0)\}, \sigma = 1$ | 3                                     |
| 4     | <b>Log normal</b> $\mu = 2 - 0.85I(Z_1 > 0.5) + 0.85Z_2, \sigma = 1$         | 4 or 2                                |

The main objective of these simulation schemes in subsection 3.1.1.3 is to gain invaluable insight into the performance of the proposed splitting rule presented in Table 3.1 and 3.2 on survival based-tree method, which is the main focus of this section. In this section, it can be concluded that Task 1 has been achieved and the various numerical results of the simulation schemes will be presented in Chapter four.

### 3.1.4 Task 2: Survival Tree-Based via between-node heterogeneity

Between node heterogeneity Classification Tree requires estimating the change in homogeneity from parent node to daughter nodes for every possible split, and later selects the outstanding split at each node of the tree. The best split is taken by measures of impurity function denoted by  $\emptyset$ . CART also uses between nodes to maximize impurity function based on selected test statistics. The value of impurity will be obtained by moving from the parent node  $p$  to the left node L and right node R.

In between node homogeneity, the main procedures of growing tree are in four phases. The first procedure requires choosing a suitable measure of homogeneity for the response (outcome). The second procedure required technique for growing the tree and by choosing appropriate stopping criteria (e.g. stopping when the sample size is very small in each terminal node or having identical covariates in tree). Thirdly, the grown tree should be pruned in order to remove the excessive trees and outliers. The final procedure requires selecting the best tree after pruning and this can be achieved by choosing the simplest tree that will categorize the observations effectively. In this study, all the four criteria will be adopted into survival model with more than one risk and subsequent paragraphs explain the adaption in full details.

In tree construction, growing and pruning are required. In this segment, we describe the methods used in both stages of constructing a tree that maximize between-node heterogeneity for data with competing risks.

#### **3.1.4.1 Growing of Survival Tree-Based via between-node heterogeneity**

The tree is grown by selecting any possible splitting function on the covariates space  $Z$  based on the event specific cumulative incidence function (CIF). The CIF is a useful quantity for competing risk problems, and is the probability of occurrence by time  $t$  for a particular type of failure in the presence of other risks. CART is a non-parametric statistics that is why CIF is used to test the equality of competing risks. Characteristically, the inference for a CIF is based on its renowned Aalen-Johansen estimator (AJE) (1978). Conversely, due to its complicated limit distribution, additional methods are needed to gain AJE-based inference methods (Dobler and Pauly, 2014). This complication can be pointed out when constructing simultaneous confidence bands for a CIF,

and can be corrected using Lin's resampling method (Lin et al 1993 and Lin 1997) or the monograph of Martinussen and Scheike (2006).

In recent times, it has been pointed out that Lin technique is a special illustration of the general Wild bootstrap, see Cai et al. (2010), Elgmati et al. (2010) and Beyersmann et al. (2013). Additionally, the weak convergence of the wild bootstrap and other weighted bootstrap versions of the AJE have been rigorously studied in Beyersmann et al. (2013) as well as Dobler and Pauly (2013 and 2014).

After growing a tree, we used the methods proposed by Gray (1988), Pepe and Mori (1991,1993), Kolmogorov-Smirnov test (2007, 2014) and Neyman Smooth test (2007) to compare the two CIFs. These techniques extended the rank-based test for survival data with univariate event (e.g. log-rank test, Generalized Wilcoxon test, Fleming-Harrington test, Tarone-Ware test and Peto-Peto-Prentice test) to apply to data with competing risks. This process (splitting) was repeated and compares the CIF until we found two groups with maximum between-node heterogeneity.

We consider  $k = 1, 2$  as an indicator of two groups, for any given split. It was known that in the competing risks model, failures or events of interest are classified into non-overlapping exclusive types. Suppose data from competing risk model are *i.i.d.* random vector  $\{(T_i, \delta_i), 1 \leq i \leq n\}$  with  $J$  different event types ( $j = 1, 2, \dots, J$ ). By assuming that the event of interest is event 1 from this point forward and events  $j = 2, \dots, J$  are referred to competing risk events.

Let  $T_{ik}^0$ ,  $C_{ik}$  and  $\delta_{ik}^0$  be the true failure time, true censoring time and indicator of true event type, respectively, for the  $i$ th individual in group  $k = 1, 2$ . Each subject  $i$  in set of data are observed as  $(T_{ik}, \delta_{ik})$ , where  $T_{ik} = \min(T_{ik}^0, C_{ik})$  and  $\delta_{ik} = \delta_{ik}^0 I(T_{ik} \leq C_{ik})$ . It was assumed that observations from different individuals within a group are independent and identically distributed;

however, we assumed that the underlying competing risk processes for each individual are not independent. Moreover, we assumed that the censoring rates are independent of the failure time  $T_{ik}^0$  for any event  $j$ .

Let  $F_{jk}(t) = P(T_{ik}^0 \leq t, \delta_{ik}^0 = j)$  denotes cumulative incidence function (CIF) for failures of type  $j$  in group  $k$ . For every possible split of a covariate node, we need to test whether the CIFs are equal or not, simply by  $H_0: F_{11} = F_{12} = F_1^0$ , where  $F_1^0$  is an unspecified sub-distribution function and where the failure type of special interest is taken to be type 1. To simplify the presentation, the  $F_1^0$  are assumed to be continuous with sub-density  $f_{jk}(t)$  with respect to Lebesgue measure. Let  $S_K(t) = P(T_{ik}^0 > t) = 1 - \sum_j F_{jk}(t)$  be the survival function for individuals in group  $k$ . In general, we take  $J = 2$  (*i.e two competing events*). It has been pointed out by Luo and Turnbull (1999) that: a subject cannot fail from two causes at the same time. For each subject  $i$ , we denote cause-specific counting processes as

$$N_{jk}(t) = \sum_{i=1}^{n_k} I(T_{ik} \leq t, \delta_{ik} = j),$$

Denote all information observed up to time  $t$  by the filtration

$$\mathcal{F}_t = \sigma\{N_{jk}(u), 1 \leq j \leq k, 1 \leq i \leq n, 0 \leq \mu \leq t\}, t \geq 0 \quad (3.13)$$

When  $k = 2$ , this filtration is consistent with the one defined in Theorem 1.3.1 of Fleming and Harrington (1991, p.26). Fleming and Harrington (1991, p.29) show that, for each  $j$ ,  $1 \leq j \leq k$ ,

$$M_{jk}(t) = N_{jk}(t) - \int_0^t \hat{\lambda}_j(s) 1_{(T \geq s)} ds, 1 \leq i \leq n$$

are locally square integrable  $F_t$ -martingales with respect to the filtration (3.13)

Where  $N_{jk}(t)$  is the number of failures of type  $j$  by  $t$  and  $Y_k(t)$  is the number of subjects still at risk just prior to  $t$  in group  $k$ . An estimate of the cumulative function is given by

$$\hat{F}_{jk}(t) = \int_0^t \hat{S}_k(u-) Y_k^{-1}(u) dN_{jk}(u), t \geq 0 \quad (3.14)$$

The first test statistics to be considered is Gray. Gray (1998) developed a class based on weighted integrals with respect to the difference of estimated cumulative sub-distribution hazard functions, defined as

$$Z = \int_0^\tau W(t) [\{1 - \hat{F}_{11}(t-)\}^{-1} d\hat{F}_{11}(t) - \{1 - \hat{F}_{12}(t-)\}^{-1} d\hat{F}_{12}(t)] \quad (3.15)$$

Where  $W(t)$  is a weight function (a predictable process) defined by Gray in 1988 as

$$W(t) = \{\hat{G}_1^0\}^\rho, -1 \leq \rho \leq 1, \text{ and}$$

$$\hat{G}_1^0 = 1 - \hat{F}_1^0(t) = 1 - n^{-1} \int_0^t \hat{h}^{-1}(u) dN_1(u),$$

Where  $n = n_1 + n_2$ ,  $\hat{h}^{-1}(t) = \frac{n^{-1} I(t \leq \tau_k) Y_k(t)}{\hat{S}_k(t-)}$ , and  $\hat{h} = \hat{h}_1 + \hat{h}_2$ .

To compare the two CIF after split, we choose to use  $\frac{Z_1^2}{n\sigma_{11}^2} \sim \chi_1^2$  asymptotically (see Gray 1988 and Fiona 2008),

Similarly, this is applied to multivariate events with the  $j^{th}$  component of the rank-based test statistic given as;

$$X_j(t) = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \int_0^\tau \hat{W}_j(s) d[\hat{F}_{j1}(t) - \hat{F}_{j2}(t)] \quad (3.16)$$

$$\hat{F}_{jk}(s) = \int_0^t \frac{\hat{S}_K(s)}{\bar{Y}_k(s)} d\bar{N}_{jk}(s)$$

$$\hat{S}_K(t) = \exp\left\{-\sum_{j=1}^J \int_0^t \frac{1}{\bar{Y}_k(s)} d\bar{N}_{jk}(s)\right\}$$

$$\bar{Y}_k(s) = \frac{1}{n_k} \sum_{i=1}^{n_k} I(T_{ik} \geq t)$$

$$\bar{N}_{jk}(s) = \frac{1}{n_k} \sum_{i=1}^{n_k} I(T_{ik} \leq t, \delta_{ik} = j)$$

and  $\hat{W}_j(s)$  is a weighted function that converges to some weight,  $\hat{W}_j(s)$ , in probability. This generated the second test statistic. The equation 3.16 is proposed by Pepe (1991) and related to the one proposed by Gray (1998). However, our interest is set on univariate inference on cumulative incidence of all causes.

Thirdly, we proposed to use Neyman Smooth test (2007) to compare the two CIFs, after growing the tree. Neyman's Smooth goodness-of-fit procedure is based on embedding the null hypothesis in a smooth alternative model described by a finite number of parameters (Kraus, 2007). The embedding is formulated in terms of densities (Rayner & Best, 1989). Testing the directional hypothesis that the data come from a distribution with density  $f = f_0$  null hypothesis is embedding into the  $d - dimensional$  alternative

$$f(x; \theta) = f_0(x) \exp\{\theta^T \varphi(F_0(x)) - c(\theta)\}, \quad x \in \mathbb{R}. \quad (3.17)$$

where  $\varphi(u) = (\varphi_1(u), \dots, \varphi_d(u))^T, u \in [0,1]$  are some square integrable basis function,  $c(\theta) = \log \int_{\mathbb{R}} f_0(x) \exp\{\theta^T \varphi(F_0(x))\} dx$  is a normalizing constant and  $F_0$  the distribution function corresponding to  $f_0$ .

The logarithm of the sub-distribution hazard ratio is expressed as a linear combination of some functions. Now  $\psi_l(t), t \in (0, \tau), l = 1, \dots, d$  are of the form  $\psi_l(t) \left( \frac{F_0(t, 1)}{F_0(\tau, 1)} \right)$ , where  $\psi_l(u), u \in [0,1]$  are some linearly independent basis functions (e.g. orthogonal Legendre Polynomials of order  $0, 1, \dots, d-1$ , or  $\sqrt{2} \cos((l-1)\pi)$ ).

Basically, in 2007, Kraus developed a score test of the hypothesis  $H_0: \theta = 0$  Vs.  $H_0: \theta \neq 0$ . He applied the Breslow-type estimator and he finally got the score vector as

$$U(\tau) = \int_0^\tau L(t) \left( d\hat{\Gamma}_2(t, 1) - d\hat{\Gamma}_1(t, 1) \right) \quad (3.18)$$

Where  $L(t) = \frac{\bar{R}_1(t, 1) \bar{R}_2(t, 1)}{\bar{R}_1(t, 1) + \bar{R}_2(t, 1)}$ .

$U(\tau)$  is a vector of weighted logrank-type statistics for comparing sub-distribution hazard functions. When  $U(\tau)$  is one-dimensional (*i.e.*  $d = 1$ ) and  $\varphi_1(t) = 1$ , it agrees with the statistic of Gray (1988). He further shows that, under the null hypothesis the score vector  $n^{1/2} U(\tau)$  (where  $n = n_1 + n_2$ ) is asymptotically normal with mean zero and variance matrix which is consistently estimated by  $n^{\frac{1}{2}} \hat{\sigma}(\tau, \tau)$  given in his theorem and shall be shown in our Appendix. Accordingly, the quadratic score statistic  $T = U(\tau)^T \hat{\sigma}(\tau, \tau)^{-1} U(\tau)$  is asymptotically  $\chi^2$  distribution with  $d$  degrees of freedom (Kraus, 2007).

The fourth test statistic is the Kolmogorov-Smirnov test that considered the problems of testing the equality of two cumulative incidence functions against a global alternative. The hypothesis is:

$$H_0: F_{11}(t) = F_{12}(t) \text{ for all } t \leq \tau$$

*Versus*

$$H_0: F_{11}(t) \neq F_{12}(t) \text{ for some } t \leq \tau$$

Where  $\tau$  is typically the largest on study time.

The test statistic was proposed by Bajorunaite et al. (2008, 2007) suggested by Lin (1997). The test statistic is based on the maximum difference between two cumulative incidence functions by the weighted version of Kolmogorov-Smirnov type;

$$T^{KS} = \sup_t W(t) |\hat{F}_{11}(t) - \hat{F}_{12}(t)|, \quad (3.19)$$

where  $W(\cdot)$  is a positive weight function. A resampling method proposed by Lin (1999) will be used to simulate the distribution of the test statistic (3.19). The derivations of the test statistic were based on the martingale representation of the process:

$$D_K(t) = \sqrt{nk} [\hat{F}_{1k}(t) - F_{1k}(t)], k = 1, 2, \quad (3.20)$$

which is the back bone of the resampling technique.  $D_K(t)$  can be rewritten as follows:

$$D_K(t) = \sqrt{n_k} \left[ \int_0^t \frac{1-F_{2k}(u)}{Y_k(u)} dM_{1K}(u) + \int_0^t \frac{F_{1k}(u)}{Y_k(u)} dM_{2K}(u) - F_{1k}(t) \left( \int_0^t \frac{dM_{1K}(u)}{Y_k(u)} + \int_0^t \frac{dM_{2K}(u)}{Y_k(u)} \right) \right] + O_p(1),$$

where  $dM_{jK}(t) = dN_{jk}(t) - Y_k(t) dA_{jK}(t)$ ,  $j=1,2$ ,  $k=1,2$ , are orthogonal martingales (Bajorunaite et al. 2007, Lin, 1997). It is very important to evaluate the process in equation (3.20), in order to make inference about  $F_{1k}(t)$ . The evaluation is difficult due to the complicated structure of the

limiting covariance function (Lin, 1997). A resampling technique will be employed. To test the null hypothesis,

$$\begin{aligned}
Q(t) &= \hat{F}_{11}(t) - \hat{F}_{12}(t) \\
&= \{\hat{F}_{1k}(t) - F_{1k}(t)\} - \{\hat{F}_{2k}(t) - F_{2k}(t)\} \\
&= \frac{D_1(t)}{\sqrt{n_1}} - \frac{D_2(t)}{\sqrt{n_2}}
\end{aligned}$$

The above resampling can be used to obtain the realizations of  $Q(t)$  and  $T^{KS}$  for any given data under the null hypothesis. This can be used to obtain a p-value for the test of equality of two cumulative incidence functions.

The last method of analysis being considered is described in Fine and Gray (1999). The methods involve analyzing competing risks in the presence of several covariates using cause-specific hazard functions to model survival data with more than one event under proportional hazard assumptions. The authors develop a semi-parametric model for the crude incidence, which is similar to the one developed by Cox (1972). The interest is on modelling the cumulative incidence function for failure due to event  $j$  (cause 1 or 2) conditional on the covariate, say,  $Z$ , the sub-distribution function is expressed in terms of sub-hazards as

$$\begin{aligned}
F_j(t; Z) &= P(T \leq t, \delta_i = 1 | Z) \\
F_j(t; Z) &= 1 - e^{\int_0^t \gamma_j(t/Z)}
\end{aligned} \tag{3.21}$$

The sub-hazard in equation 3.21 was fitted with a Cox model and gives:

$$\gamma_j(t/Z) = \gamma_{0j}(t) e^{\beta_j' Z}$$

Fine and gray (1999) showed that the partial likelihood valid for estimation is given by

$$\tilde{L}(\beta_j) = \prod_{i=1}^n \left( \frac{e^{\beta_j' z_i}}{\sum_{l \in \hat{R}_i} w_{il} e^{\beta_l' z_i}} \right)$$

This implies, the risk for cause  $j$  at time  $t_i$  is  $\hat{R}_i = \{l: t_l \geq t_i \text{ or } (t_l \leq t_i \text{ and } C \neq j)\}$  and the subjects experiencing a competing cause remain in the risk set. The  $w_{il}$  is given to any individual as  $w_{il} = \frac{\hat{G}(t_i)}{\hat{G}(\min(t_l, t_i))}$ , where  $\hat{G}$  is the survivor function for the censoring distribution.

### 3.1.4.2 Pruning of Survival Tree-Based via between-node heterogeneity

In order to adjust for over fitting on the impurity function  $\phi(h)$  described above, the tree grown has to be pruned. We use the cost complexity algorithm presented by Breiman et al. (1984) to obtain an ordered list of optimally pruned sub-trees, substituting our impurity function for node  $h$ ,  $\phi(h)$ , based on event-specific residuals LeBlanc and Crowley (1992,1993).

Cost-complexity function of tree  $T$  is given as

$$\phi_\alpha(T) = \sum_{h \in T^t} \phi(h) + \alpha |T^t|, \quad (3.22)$$

where  $\alpha$  is a nonnegative complexity parameter.

The tree that minimizes the cost-complexity function (3.22) is the optimally pruned sub-tree. As  $\alpha$  increases, there are “jumps” whereby a smaller sub-tree becomes the minimizing tree for a new value of  $\alpha$ , say  $\alpha^*$ . The  $\alpha^*$ ’s are defined as

$$\alpha^* = \frac{\phi(h) - \phi(T_h)}{|T_h^t| - 1} \quad (3.23)$$

where  $\phi(T) = \sum_{h \in T^t} \phi(h)$  and  $T_h$  represent the root node  $h$  of the tree. The subdivision with the smallest cost complexity is pruned recursively until we have a decreasing sequence of smallest optimally pruned sub-trees.

$$T_m < T_{m-1} < \dots < T_1 < T_0$$

and the arrangement of corresponding  $\alpha^*$ 's is

$$\infty = \alpha_m^* > \alpha_{m-1}^* > \dots > \alpha_1^* > \alpha_0^* = 0$$

In order to select the final tree, cross validation method was adopted and used to grow an optimally pruned tree through the sequence of trees  $T_m < T_{m-1} < \dots < T_1 < T_0$  on the entire data  $\mathcal{L}$ . Then, the data set was divided as  $\mathcal{L}_{(v)} = \mathcal{L} - \mathcal{L}_v$ . For each  $v$ , we use  $\mathcal{L}_{(v)}$  to grow a nested sequence of optimally pruned trees.

$$T_m^v < T_{m-1}^v < \dots < T_1^v < T_0^v, \text{ which we obtain by pruning } T_0^v \text{ using } \alpha'_m = \sqrt{(\alpha_m^* \alpha_{m+1}^*)}.$$

For the data set  $\mathcal{L}_v$ , we calculate the sum of statistic for between node statistics  $G(T_m^v)$ . To estimate  $G(T_m)$ , we average the sum of over  $V$  set as:

$$G^{cv}(T_m) = \frac{1}{V} \sum_{i=1}^V G(T_m^v), \quad (3.24)$$

and choose the  $\alpha = \alpha^*$  that maximizes this average (equation 3.23).

Our final tree,  $T(\alpha_c)$ , is chosen from the original list of optimally pruned sub-trees grown on the whole sample  $\mathcal{L}$ . The  $T(\alpha_c)$  is the smallest optimally pruned sub-tree that has the minimum cross-validation adjusted  $\phi_\alpha(T)$  for  $\alpha = \alpha_c$ . This means it is

$$T(\alpha_c) = \arg_{T_m \in \{T_0, \dots, T_m\}} \phi^{cv}(T_m) + \alpha_c |T_m^t|$$

### 3.1.4.3 Simulation Scheme for Survival Tree-Based via between-node heterogeneity

In this section, this study simulates data to investigate how accurately the proposed prognostic survival based-tree method detects cut points based on the covariates. This was done by simulating data with two competing risks. Table 3.6 revealed the procedures and model modality adopted to carry out the simulations. There are two models in the table, each of them was spawned from both Exponential and Lognormal distributions with 50% censoring rate. 200 trees were constructed with 20, 50, 80, 100, 150, 200, 500 and 1000 sample sizes for each simulation. All models' results were presented in this section for both exponential and log-normal distributions. The numerical results and its explanations were given in Chapter four.

Table 3. 6: Simulation Procedures of Classification Tree Based on Between Node Heterogeneity

| Exponential distribution |                                                |                 |                                             |                  |
|--------------------------|------------------------------------------------|-----------------|---------------------------------------------|------------------|
|                          | Cumulative Incidence Function (CIF)<br>Event 1 |                 | Cumulative Incidence Function (CIF) Event 2 |                  |
|                          | <i>Moderate censoring rate (50%)</i>           |                 |                                             |                  |
| Cut-point                | $F_1(t Z \leq c)$                              | $F_1(t Z > c)$  | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$   |
| 0.5                      | Exp(0.2)                                       | Exp(0.4)        | Exp(0.7)                                    | Exp(0.5)         |
| Lognormal Distribution   |                                                |                 |                                             |                  |
| Cut-point                | $F_1(t Z \leq c)$                              | $F_1(t Z > c)$  | $F_1(t Z \leq c)$                           | $F_1(t Z > c)$   |
| 0.5                      | Log-normal(1.25)                               | Log-normal(0.5) | Log-normal(0.1)                             | Log-normal(0.25) |

## 3.2 Task 3: Extension of Artificial Neural Network in Survival Analysis with Competing Risks

The original function of neural network (NN) was to use computer-based model to mimic human brain. The most widely used NN is called the Multi-layer perceptron (MLP) system, which

consists of a set of input features, a number of interactive layers, in between the inputs and output layers and specifically one output layer (Yahya et al; 2012). Fig. 1.3 & 1.4 in chapter one demonstrated the functional relationship between the two input variables input 1 and input 2 and the response (output) variable while the appropriate network parameters (weight) were estimated. In addition, the activated function ( $F(X)$ ) is used to transform the processed input into output (Paula and Joao., 2008).

The simplest multilayers perceptron is given in the equation below

$$o(x) = f(w_0 + \sum_{i=1}^n w_i x_i) , \quad (3.25)$$

where  $w_0$  denotes the intercept,  $w_i = (w_1, w_2, \dots, w_n)$  the vectors consisting of all synaptic weight without the intercept,  $x = (x_1, x_2, \dots, x_n)$  is the vector of all covariates and  $o(x)$  is the response variable (Frauk et al., 2010).

The study proposed to use the event-specific Cox-Snell residual as an outcome measure for event  $m$  risk ( $m = 1$  and  $2$ ), for individual  $i$  in group  $k$ , this residual for the event of interest ( $j = m$ ) is defined as

$$\hat{r}_{c_{ik}}^m = I_{\{\delta_{ik}=m\}} - \hat{\Lambda}_0^m(T_{ik}) \quad (3.26)$$

Where  $\hat{\Lambda}_0^m(.)$  is the estimated cause-specific cumulative hazard function. This function can be estimated by

$$\hat{\Lambda}_0^m(T_{ik}) = \int_0^{T_{ik}} \frac{dN_{1k}(S)}{Y_k(S)}$$

where  $N_{1k}(t)$  and  $Y_k(t)$  are the number of event 1 and number of risks at time  $t$  for group  $k$ , respectively (Dauda et al., 2015).

Moreover, we proceed to formulate a new neural network by substituting the response variable in equation 3.26 by equation 3.25. Thus, we have our model to be

$$\exp(\beta'x_i) \hat{H}_0(t_i) = f\left(w_0 + \sum_{i=1}^n w_i x_i\right).$$

Hence, the estimate of our proposed model by equation 4.4 becomes

$$\hat{r}_{cik}^m = f(w_0 + \sum_{i=1}^n w_i x_i), \text{ for all } i = 1, 2, \dots, n.$$

Additionally, we proposed to increase the model flexibility by the inclusion of hidden layers. Conversely, it has been prove by Hornik et al. (1989) that one hidden layer is sufficient to model any continuous functions. However, the Cox Multi-layer perceptron (CMLP) with J hidden neuron layers can be estimated as follows:

$$\hat{r}_{cik}^m = f\left[w_0 + \sum_{j=1}^J w_j \cdot f\left(w_{0j} + \sum_{i=1}^n w_{ij} x_i\right)\right]$$

where  $w_0$  denotes the intercept of the output neuron and  $w_{0j}$  the intercept of the  $j^{th}$  hidden neuron. Likewise,  $w_j$  denotes the synaptic weight corresponding to the synapse starting at the  $j^{th}$  hidden neuron and leading to the output neuron,  $\mathbf{w}_j = (w_{1j}, w_{2j}, \dots, w_{nj})$  the vector of all synaptic weights corresponding to the synapses leading to the  $j^{th}$  hidden neuron, and  $x = (x_1, x_2, \dots, x_n)$  the vector of all covariates. It has been clearly stated by Frauke and Stefan. (2010) that the weight is the parameter of neural network and in the proposed model, weight is also considered as parameters to be estimated. All weights are usually initialized with random values drawn from a standard normal distribution. During an iterative training process, the following steps are repeated:

- a. The neural network calculates an output  $\hat{r}_{cik}^m$  for given inputs  $x$  and current weights. If the training process is not yet completed, the predicted output  $\hat{r}_{cik}^{*m}$  will differ from the observed output  $\hat{r}_{cik}^m$ .

- b. An error function  $E$ , the sum of squared errors (SSE)

$$E = \frac{\sum_{l=1}^L \sum_{h=1}^H [(\hat{r}_{c_{ik}}^{*m})_{lh} - (\hat{r}_{c_{ik}}^m)_{lh}]^2}{2}.$$

The cross-entropy is  $E^{ce}$

$$E^{ce} = - \sum_{l=1}^L \sum_{h=1}^H \left[ \left( (\hat{r}_{c_{ik}}^m)_{lh} \log \left( (\hat{r}_{c_{ik}}^{*m})_{lh} \right) + (1 - (\hat{r}_{c_{ik}}^m)_{lh}) \log(1 - (\hat{r}_{c_{ik}}^{*m})_{lh}) \right) \right]$$

The Error measures the difference between predicted and observed output, where  $l = 1, \dots, L$  indexes the observations, i.e. given input-output pairs, and  $h = 1, \dots, H$  the output nodes.

- c. All weights are adapted according to the rule of a learning algorithm. These procedures are similar to Frauke and Stefan (2010) (see Appendix for the R codes).
- d. Finally, we adopted Resilient back propagation algorithm to train the ANN in the model CMLP. Also, four different hidden (2, 5, 10 and 15) layers were applied with sigmoid transfer function.

### 3.2.1 Relative Important by Garson Algorithm

After experimenting with the proposed model CMLP on the real-life data, to measure the contribution of independent (input) variables on the response variable, the Garson Algorithm (1991) were employed. The algorithm partitions the hidden layer weights into components associated with each input node. Subsequently, the percentage of all hidden node weight associated with the input node was used to measure the relative importance of that attribute. In general, each input  $j$ , where  $j = 1, 2, \dots, i$ , the relative importance ( $RI_j$ ) can be calculated using the equation (5).

$$RI_j = \frac{\sum_{m=1}^{N_h} \left[ \frac{|w_{jm}^{ih}|}{\sum_{k=1}^{N_i} |w_{km}^{ih}|} \times |w_{jm}^{ho}| \right]}{\sum_{n=1}^{N_i} \left[ \sum_{m=1}^{N_h} \left[ \frac{|w_{nm}^{ih}|}{\sum_{k=1}^{N_i} |w_{kw}^{ih}|} \times |w_{jm}^{ho}| \right] \right]} \quad (3.27)$$

In equation (5),  $N_i$  and  $N_h$  are the number of input and hidden neuron respectively,  $w$  is the connection weight, the superscripts "i", "h" and "o" refer to input, hidden and output layers, respectively and subscripts "k", "m" and "n" refer to input, hidden and input neurons used. In our case there is only one output neuron. In this algorithm, each input node  $j$ , the relative contribution of  $j$  to the outgoing signal of each hidden neuron is calculated and presented in percentage, which then serve as measure of importance for each node of the given variable. Additionally, the entire input variables with smallest contribution to the final output of the network is eliminated (Garson, 1991).

To validate the performance of the proposed model survival data with competing risk, three real-life data sets were considered. These datasets were analyzed based on the proposed technique and their empirical results were compared to the classical survival model. The various results are presented in Chapter four.

### 3.3 Task 4: Extension of Multivariate Adaptive Regression Splines in Survival Analysis

The MARS techniques has been so far actualized and utilized for continuous and binary responses. The aim of this work is to extend this method to survival time information in which the Cox-Snell residuals of a Cox PH model shall be used as response variable in a typical MARS approach.

In spite of the semi-parametric nature of the Cox PH model which makes it different from the standard parametric model with less assumption on the survival time, there exists a number of diagnostic tests that can be employed to check the model's adequacy, misspecification, outliers and influential points just to mention a few. These diagnostic tests include the test based on re-estimation (the Link test), the tests based on residuals (Schoenfeld Residuals, Cox-Snell Residuals, Martingale Residuals and Deviance Residuals) and the graphical diagnostic methods (see Hess, 1995 and Garrett, 1997 for more detail).

Of the above residual types of the Cox PH model, the Martingale and Deviance residuals have been successfully employed as response variables in the common MARS implementations (Kriner, 2007). In this work therefore, a model that used the Cox-Snell Residual for survival analysis with MARS is proposed. Results from this new model were compared with those that employed the Martingale residuals with MARS technique.

Basically, the Cox-Snell (CS) residual like its other counterparts is often employed as a measure of goodness-of-fit of the Cox PH model. Under the generalized residual settings, the Cox PH model using the survivor function

$$S(t; X) = [S_0(t)]^{\exp(\beta X)}$$

is fitted to the data. In terms of hazard, the model fitted is of the form

$$h(t, X) = h_0(t)\exp(\beta X)$$

Therefore, for each person with covariates  $X$ , the estimated survivor function is

$$\hat{S}(t)(t, X) = [\hat{S}_0(t)]^{\exp(\beta X)} \quad (3.28)$$

This gives a predicted survival probability at each time  $t$  in the dataset.

Then, the CSR is calculated by

$$\hat{\Lambda}_j = -\log[\hat{S}(T_j; X)] \quad (3.29)$$

Additionally, CS residual can be used to examine the overall fit of the Cox PH model by computing

$$h_j(t) = \hat{H}_0(t) \exp\left(\sum_{k=1}^p b_k x_{jk}\right), j = 1, \dots, n$$

where  $b_k = (b_1, \dots, b_p)^t$  is the Maximum Likelihood Estimator (MLE) of  $\beta$  and  $\hat{H}_0(t)$  is the MLE of the baseline cumulative hazard function (Klein et al. 2003).

Within the framework of MARS technique, the whole space of the input variable is divided into  $k$  sub-regions or groups. MARS defines a different mathematical equation for each group. This equation relates each sub-region of input variable to the output variable.

The CSR is then computed using the following quantity;

$$\widehat{CS}_{ik}^1 = I_{\{\delta_{ik}=1\}} - \hat{\Lambda}_0^1(T_{ik}) \quad (3.30)$$

where  $\hat{\Lambda}_0^1(.)$  is the estimated cause-specific cumulative hazard function which is computed by

$$\hat{\Lambda}_0^1(T_{ik}) = \int_0^{T_{ik}} \frac{dN_{1k}(s)}{Y_k(s)} \quad (3.31)$$

where  $N_{1k}(t)$  and  $Y_k(t)$  are the number of the event observed and the number of subjects at risks at time  $t$  in group  $k$  respectively.

MARS uses the following two-sided truncated power functions as splines basis functions:

$$[-(x - t)]_+^q = \begin{cases} (t - x)^q, & \text{if } x < t \\ 0, & \text{otherwise} \end{cases}$$

$$[+(x - t)]_+^q = \begin{cases} (x - t)^q, & \text{if } x \geq t \\ 0, & \text{otherwise} \end{cases}$$

where  $q$  is the power and  $t$  is the number of knots. This gives rise to the final model

$$\hat{Y} = \hat{f}(x) = a_0 + \sum_{m=1}^m a_m B_m(x) \quad (3.32)$$

where  $\hat{Y}$  is the output variable,  $x$  is the input variable,  $a_0$  is the constant term,  $m$  is the number of the splines functions, and  $a_m$  and  $B_m$  are the  $m^{th}$  splines function and its coefficient respectively.

The execution of the MARS technique follows the algorithm explained in Chapter two and the procedures still holds for survival analysis

### 3.3.1 Simulation Studies of the Survival MARS

Simulation studies are an important statistical tool often employed to examine the performance, power and other properties of a new statistical method or approach. Thus, various schemes employed for the Monte-Carlo experiment in this study are presented in this section.

#### i. Simulating Survival Times

Survival times were simulated from the three selected distributions; i.e. Exponential, Weibull and Gompertz distributions. Table 3.17 shows the methods employed for generating the survival times across these three distributions. Some general considerations employed in the Monte-Carlo experiment are based on the following:

The survival function of the Cox PH model is given

$$S(t/x) = \exp[-H_0(t)\exp(f(x))]$$

and the cumulative survival function is defined as

$$F(t/x) = 1 - S(t/x)$$

where  $H_0(t) = \int_0^t h_0(u)du$ , is the cumulative baseline hazard function (Yahya and Ulm, 2009).

Table 3. 7: Statistics for generating survival times across the specified distributions

|                        | Distribution                              |                                                                        |                                                                                          |
|------------------------|-------------------------------------------|------------------------------------------------------------------------|------------------------------------------------------------------------------------------|
|                        | Exponential                               | Weibull                                                                | Gompertz                                                                                 |
| <b>Parameter</b>       | Scale parameter $\lambda > 0$             | Scale parameter $\lambda > 0$<br>Shape parameter $V > 0$               | Scale parameter $\lambda > 0$<br>Shape parameter $\alpha \in (-\infty, \infty)$          |
| <b>Survival time</b>   | $T = -\frac{\log(u)}{\lambda \exp(f(x))}$ | $T = \left[ -\frac{\log(u)}{\lambda \exp(f(x))} \right]^{\frac{1}{v}}$ | $T = \frac{1}{\alpha} \log \left[ 1 - \frac{\alpha \log(u)}{\lambda \exp(f(x))} \right]$ |
| <b>Hazard function</b> | $h(t) = \lambda \exp(f(x))$               | $h(t) = \lambda \exp(f(x)) v t^{v-1}$                                  | $h(t) = \lambda \exp(f(x)) \exp(\alpha t)$                                               |

Five covariates  $x_1, \dots, x_5$  were simulated over  $n = 1000$  samples from different distributions as follows:  $x_1 \sim U[30, 75]$ ,  $x_2 \sim \text{bin}(n, 0.5)$ ,  $x_3 \sim N(0, 1)$ ,  $x_4 \sim U[0, 10]$  and  $x_5 \sim \text{bin}(n, 0.7)$  with  $u \sim U[0, 1]$ .

In the simulation of survival times ( $t$ ) under the three distributions (Exponential, Weibull and Gompertz) employed in this work, only the linear, nonlinear and interaction effects of covariates  $x_1$  and  $x_2$  on the response variables ( $t$ ) were considered. The other covariates  $x_3, x_4$  and  $x_5$  were introduced into the model as nuisance covariates that are expected to be identified by the survival MARS model itself. This concept was first introduced by Friedman (1991) and later by Kriner (2007).

Therefore, for the Exponential distribution, the survival time  $t$  was simulated according to the scheme adopted by Kriner (2007) using the relation:

$$t = \frac{-\log(u)}{0.5 \exp\left[0.04\left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right]}$$

For the Weibull distribution, the survival time  $t$  was simulated from the relation

$$t = \left[ \frac{-\log(u)}{\lambda \exp\left[0.04\left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right]} \right]^{\frac{1}{\nu}}$$

with the shape parameter  $\nu = 2$  and the scale parameter  $\lambda = 5$  (See Kriner, 2007).

For the Gompertz distribution, the density and survival functions of this extreme value distribution with parameters  $\lambda$  and  $\alpha$  are given by

$$f_0(t) = \lambda \exp(\alpha t) \exp\left(-\frac{\lambda}{\alpha} \exp(\alpha t)\right) \quad (3.33)$$

$$S_0(t) = \exp\left(-\frac{\lambda}{\alpha} \exp(\alpha t)\right) \quad (3.34)$$

with the expressions for the mean and variance given by;

$$E(T) = \mu_0 = -\frac{1}{\alpha} \left( \log\left(\frac{\lambda}{\alpha}\right) + \gamma \right) \quad (3.35)$$

$$V(T) = \sigma_0^2 = \frac{\pi^2}{6\alpha^2} \quad (3.36)$$

where  $\gamma \approx 0.5772$  is the Euler's constant and  $\pi = 3.14159$ . Solving equation (3.35) and (3.36) for  $\lambda$  and  $\alpha$  gives  $\alpha = \frac{\pi}{\sqrt{6}\sigma_0}$ ,  $\lambda = \alpha \exp(-\gamma - \alpha\mu_0)$ .

The values of  $\alpha$ ,  $\lambda$  and  $\gamma$  determined above were used to compute the approximate parameters of the Gompertz distribution independence on the mean and variance of the considered survival time. The adopted scheme is similar to that of Bender et al. (2005) by setting the mean life expectancy  $\mu_0$  at  $\mu_0 = 66.86$  years with a standard deviation of  $\sigma_0 = 6$  years. By this, the values of  $\lambda$  and  $\alpha$  were explicitly determined to be  $\lambda = 7 \times 10^{-8}$  and  $\alpha = 0.2138$ . Hence, the survival time  $t$  was simulated from the relation

$$t = \frac{1}{0.2138} \log \left[ 1 - \frac{0.2138 \log(u)}{0.0000007 \exp\left(0.04 \left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right)} \right]$$

### 3.4 Summary of Chapter Three

This chapter starts with a brief explanation of the methodology to follow in the upcoming sections and subsections. Broadly speaking, the methodological approach is based on the data mining approaches, these techniques were successfully extended to survival analysis with and without competing risks. The three data mining adopted (i.e. CART, ANN and MARS) were based on the critical literature review and with reference to the drawback of classical survival methodology highlighted in the previous chapters.

This chapter also described the simulation scheme used on this research. In particular, it emphasized the appropriate survival time and status to use, the characteristics of the datasets and the most suitable survival based-tree on both data structure and cutpoint.

Furthermore, this chapter stated how the methodology building process is divided into four main tasks: Classification Regression Tree (CART) via within-node homogeneity at the first step (Task 1) and between-node heterogeneity at the second step (Task 2), Multi-layer Perceptron Artificial

Neural Network (MLPANN) at the third step (Task 3) and Multivariate Adaptive Regression Spline (MARS) at the last step (Task 4).

Finally, the next chapter (Chapter four) provides the simulation results for the various simulation schemes highlighted in this chapter, the empirical results for various comparisons were also showcased and the various published real-life data results were generated and presented appropriately.

## **CHAPTER FOUR**

### **DISCUSSION OF RESULTS**

This chapter discusses the results of the four-task methodology applied in this study: Classification Regression Tree (CART) via within-node homogeneity survival based-tree (Task 1) and between-node heterogeneity survival based-tree (Task 2), Multi-layer Perceptron Artificial Neural Network in survival data (MLPANN) (Task 3) and Multivariate Adaptive Regression Spline (MARS) in survival data (Task 4). These results were generated based on the different simulations schemes explained in chapter three and some published real-life data were applied to the designed and extended methodologies.

#### **4.1 Survival Tree-Based via within-node homogeneity (Task 1) Numerical Results**

In this section, the study discusses a series of simulations designed to investigate how accurately the splitting rule methods detect the stipulated cutpoints and how well the tree procedures detect the data structure.

#### 4.1.1 Cutpoint detection results on survival based-tree via within-node homogeneity

Simulation studies were conducted on models in Table 3.3, which was adopted to assess the performance of the proposed model in terms of cutpoint determination, under exponential distribution. The splitting rule simulation results based on the within-node homogeneity were presented in Tables 4.1, 4.2 and 4.3. Some abbreviations were used in the Tables and the full meanings have been pointed out in chapter one under definition of terms. Meanwhile here are the few ones used in this subsection: MRSS= martingale residual sum-of-square, MRAV= martingale residual absolute value, DRSS= deviance residual sum-of-square, DRAV= deviance residual absolute values, CSRSS= Cox-Snell residual sum-of-square and CSRS= Cox-Snell residual absolute value.

The performance of the proposed model based on mean and median over the existing methodology adopted by Fiona (2008) was revealed. As a matter of fact, the proposed model estimated values that were much closer to the cut point selected than the existing techniques established by exponential distribution irrespective of the censoring rate (i.e. whether low, moderate or high), for example, in Table 4.1 the mean and median of MRSS are 0.4076 and 0.3147 while DRSS(Proposed) are 0.3521 and 0.2277. The DRSS(Proposed) mean and median are closer to the cutpoint 0.1 than the MRSS and therefore being regarded as the best method, and other comparisons were made in similar manner.

##### i. Results of Exponential Distribution over the various censoring rates (25%, 50% & 75%)

Table 4.1: Low censoring rate 25%

|                | <b>Cutpoint <math>c = 0.1</math></b> |               |               |               |               |
|----------------|--------------------------------------|---------------|---------------|---------------|---------------|
|                | Mean                                 | Median        | SD            | RMSE          | MAE           |
| MRSS           | 0.4076                               | 0.3147        | 0.2851        | 0.4194        | 0.3090        |
| MRAV           | 0.4799                               | 0.4851        | <b>0.2132</b> | 0.4356        | 0.3801        |
| DRSS(Proposed) | <b>0.3521</b>                        | <b>0.2277</b> | 0.2656        | <b>0.3661</b> | <b>0.2529</b> |

|                         |               |               |               |               |               |
|-------------------------|---------------|---------------|---------------|---------------|---------------|
| DRAV(Proposed)          | 0.4351        | 0.4151        | 0.2656        | 0.3991        | 0.3351        |
| CSRSS(Proposed)         | 0.4001        | 0.2993        | 0.2821        | 0.4119        | 0.3013        |
| CSRAV(Proposed)         | 0.4593        | 0.4531        | 0.2160        | 0.4192        | 0.3595        |
| <b>Cutpoint c= 0.5</b>  |               |               |               |               |               |
| MRSS                    | 0.5362        | 0.5166        | 0.1683        | 0.1721        | 0.1232        |
| MRAV                    | 0.5373        | 0.5175        | 0.1034        | 0.1099        | 0.0758        |
| DRSS(Proposed)          | 0.5526        | 0.5206        | 0.1293        | 0.1396        | 0.0900        |
| DRAV(Proposed)          | 0.5377        | 0.5189        | <b>0.0835</b> | <b>0.0916</b> | <b>0.0595</b> |
| CSRSS(Proposed)         | 0.5331        | <b>0.5166</b> | 0.1674        | 0.1706        | 0.1222        |
| CSRAV(Proposed)         | <b>0.5328</b> | 0.5168        | 0.1030        | 0.1081        | 0.0743        |
| <b>Cut point c= 0.9</b> |               |               |               |               |               |
| MRSS                    | 0.6739        | 0.8226        | 0.2699        | 0.3520        | 0.2290        |
| MRAV                    | 0.6738        | 0.7409        | 0.2189        | 0.3148        | 0.5739        |
| DRSS(Proposed)          | <b>0.7669</b> | <b>0.8737</b> | 0.2245        | <b>0.2609</b> | <b>0.1376</b> |
| DRAV(Proposed)          | 0.7382        | 0.8350        | <b>0.2092</b> | 0.2644        | 0.6384        |
| CSRSS(Proposed)         | 0.6702        | 0.8191        | 0.2747        | 0.3581        | 0.2330        |
| CSRAV(Proposed)         | 0.6851        | 0.7595        | 0.2183        | 0.3063        | 0.5851        |

Table 4.2: Moderate censoring rate 50%

|                         |               |               |               |               |               |
|-------------------------|---------------|---------------|---------------|---------------|---------------|
| <b>Cut point c= 0.1</b> |               |               |               |               |               |
|                         | <b>Mean</b>   | <b>Median</b> | <b>SD</b>     | <b>RMSE</b>   | <b>MAE</b>    |
| MRSS                    | 0.4206        | <b>0.3284</b> | 0.2845        | 0.4286        | <b>0.3218</b> |
| MRAV                    | 0.5057        | 0.5016        | 0.2110        | 0.4572        | 0.4058        |
| DRSS(Proposed)          | <b>0.4237</b> | 0.3347        | 0.2853        | 0.4314        | 0.3248        |
| DRAV(Proposed)          | 0.4564        | 0.4406        | 0.2226        | <b>0.4201</b> | 0.3566        |
| CSRSS(Proposed)         | 0.4289        | 0.3651        | 0.2819        | 0.4331        | 0.3302        |
| CSRAV(Proposed)         | 0.5179        | 0.5269        | <b>0.2074</b> | 0.4665        | 0.4179        |
| <b>Cut point c= 0.5</b> |               |               |               |               |               |
| MRSS                    | 0.5420        | 0.5210        | 0.1850        | 0.1897        | 0.1405        |
| MRAV                    | 0.5591        | 0.5358        | <b>0.1178</b> | 0.1318        | 0.0943        |
| DRSS(Proposed)          | 0.5429        | 0.5213        | 0.1823        | 0.1872        | 0.1385        |
| DRAV(Proposed)          | <b>0.5281</b> | <b>0.5160</b> | 0.1247        | <b>0.1278</b> | <b>0.0899</b> |
| CSRSS(Proposed)         | 0.5413        | 0.5214        | 0.1813        | 0.1859        | 0.1374        |
| CSRAV(Proposed)         | 0.5575        | 0.5350        | 0.1182        | 0.1314        | 0.0933        |
| <b>Cutpoint c= 0.9</b>  |               |               |               |               |               |
| MRSS                    | 0.6622        | 0.8047        | 0.2750        | 0.3635        | <b>0.2409</b> |
| MRAV                    | 0.6954        | 0.7776        | <b>0.2254</b> | <b>0.3044</b> | 0.5956        |
| DRSS(Proposed)          | 0.6588        | <b>0.8053</b> | 0.2777        | 0.3677        | 0.2444        |
| DRAV(Proposed)          | 0.6353        | 0.6912        | 0.2296        | 0.3503        | 0.5354        |
| CSRSS(Proposed)         | 0.6535        | 0.8003        | 0.2807        | 0.3735        | 0.2499        |
| CSRAV(Proposed)         | <b>0.6978</b> | 0.7911        | 0.2287        | 0.3053        | 0.5979        |

Table 4.3: High censoring rate 75%

|                        |             |               |           |             |            |
|------------------------|-------------|---------------|-----------|-------------|------------|
| <b>Cutpoint c= 0.1</b> |             |               |           |             |            |
|                        | <b>Mean</b> | <b>Median</b> | <b>SD</b> | <b>RMSE</b> | <b>MAE</b> |
| MRSS                   | 0.4383      | 0.3773        | 0.2824    | 0.4406      | 0.3390     |

|                        |               |               |               |               |               |
|------------------------|---------------|---------------|---------------|---------------|---------------|
| MRAV                   | 0.5286        | 0.5415        | <b>0.2009</b> | 0.4734        | 0.4287        |
| DRSS(Proposed)         | <b>0.4357</b> | <b>0.3751</b> | 0.2817        | 0.4382        | <b>0.3365</b> |
| DRAV(Proposed)         | 0.4685        | 0.4612        | 0.2129        | <b>0.4256</b> | 0.3686        |
| CSRSS(Proposed)        | 0.4429        | 0.3882        | 0.2848        | 0.4457        | 0.3438        |
| CSRAV(Proposed)        | 0.5391        | 0.5395        | 0.2027        | 0.4836        | 0.4391        |
| <b>Cutpoint c= 0.5</b> |               |               |               |               |               |
| MRSS                   | 0.5544        | 0.5331        | 0.1937        | 0.2011        | 0.1535        |
| MRAV                   | 0.5814        | 0.5593        | 0.1320        | 0.1550        | 0.1144        |
| DRSS(Proposed)         | 0.5522        | 0.5304        | 0.1942        | 0.2010        | 0.1538        |
| DRAV(Proposed)         | <b>0.5403</b> | <b>0.5247</b> | 0.1303        | <b>0.1364</b> | <b>0.0983</b> |
| CSRSS(Proposed)        | 0.5558        | 0.5348        | 0.1900        | 0.1980        | 0.1507        |
| CSRAV(Proposed)        | 0.5832        | 0.5626        | <b>0.1299</b> | 0.1542        | 0.1149        |
| <b>Cutpoint c= 0.9</b> |               |               |               |               |               |
| MRSS                   | 0.6369        | 0.7688        | 0.2799        | 0.3841        | 0.2662        |
| MRAV                   | <b>0.6944</b> | 0.7941        | 0.2400        | <b>0.3160</b> | 0.5947        |
| DRSS(Proposed)         | 0.6316        | 0.7589        | 0.2812        | 0.3887        | 0.2714        |
| DRAV(Proposed)         | 0.6342        | 0.6966        | <b>0.2355</b> | 0.3551        | 0.5344        |
| CSRSS(Proposed)        | 0.6402        | 0.7770        | 0.2833        | 0.3844        | <b>0.2631</b> |
| CSRAV(Proposed)        | 0.6936        | <b>0.7970</b> | 0.2455        | 0.3207        | 0.5938        |

The simulation results under exponential distribution were further presented using: Root Mean Square Errors (RMSE) and Mean Absolute Errors (MAE) as shown in figs. 4.1, 4.2 and 4.3. The various values of selected cut points varying from 0.1 to 0.9 were plotted against RMSE and MAE. This was done to evaluate the performances and the efficiency of the proposed models based on different levels of censoring rates. The figures showcase the efficiency of the proposed model over the existing ones since the blue line that represents the proposed model was found below the existing ones under low censoring rate irrespective of the impurity function. Nonetheless, the existing model performs equally to the proposed model when the censoring rates were moderate and yet when the censoring rate is high the proposed model outperforms the existing techniques.

Finally, the boxplot virtual display in Fig.4.4, distinguishes between the proposed model and the existing model and it also reveals the ability of the proposed model to detect cut point more efficiently than the existing model. In some cases, we also observed that the existing techniques out-fit the proposed model only when the censoring rates is low with smaller cut point.

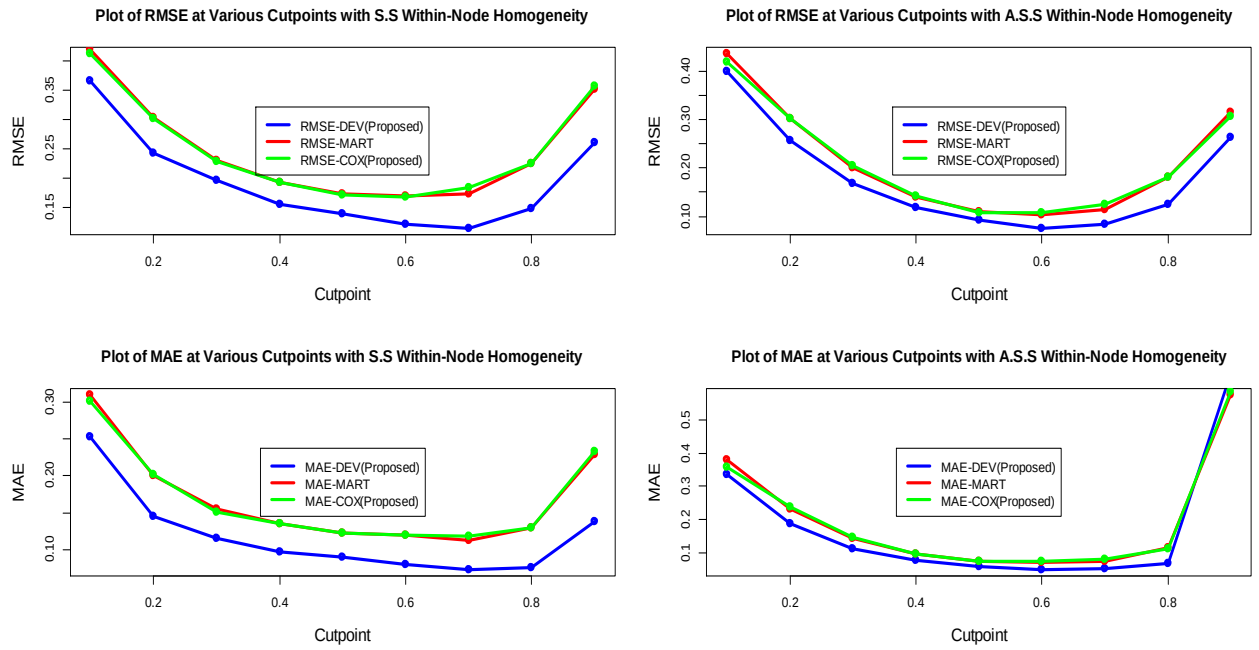


Fig. 4. 1: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 25% censoring rate (Exponential Distribution)

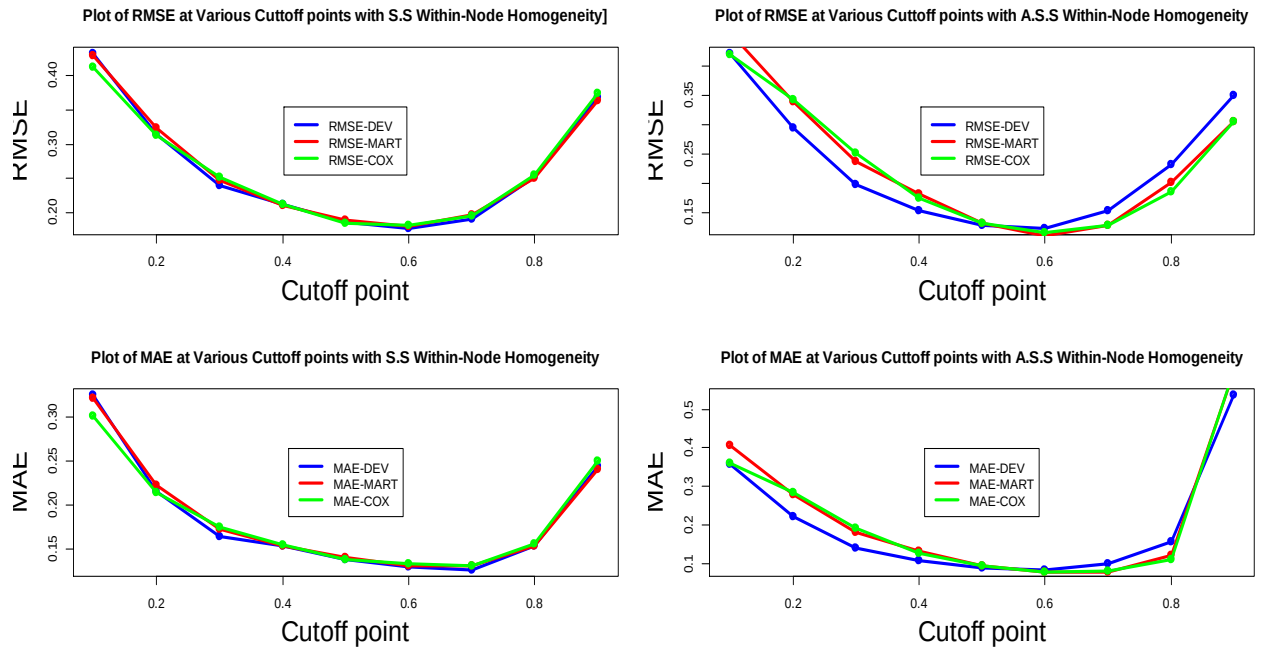


Fig. 4.2: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 50% censoring rate (Exponential Distribution)

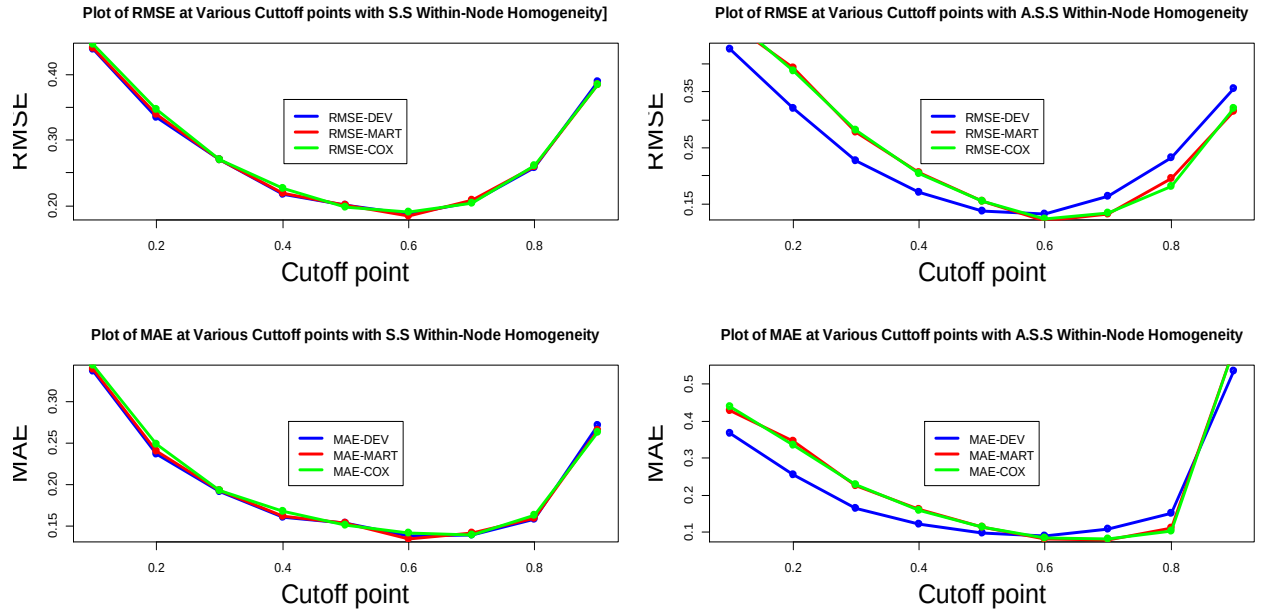


Fig. 4.3: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 75% censoring rate (Exponential Distribution)

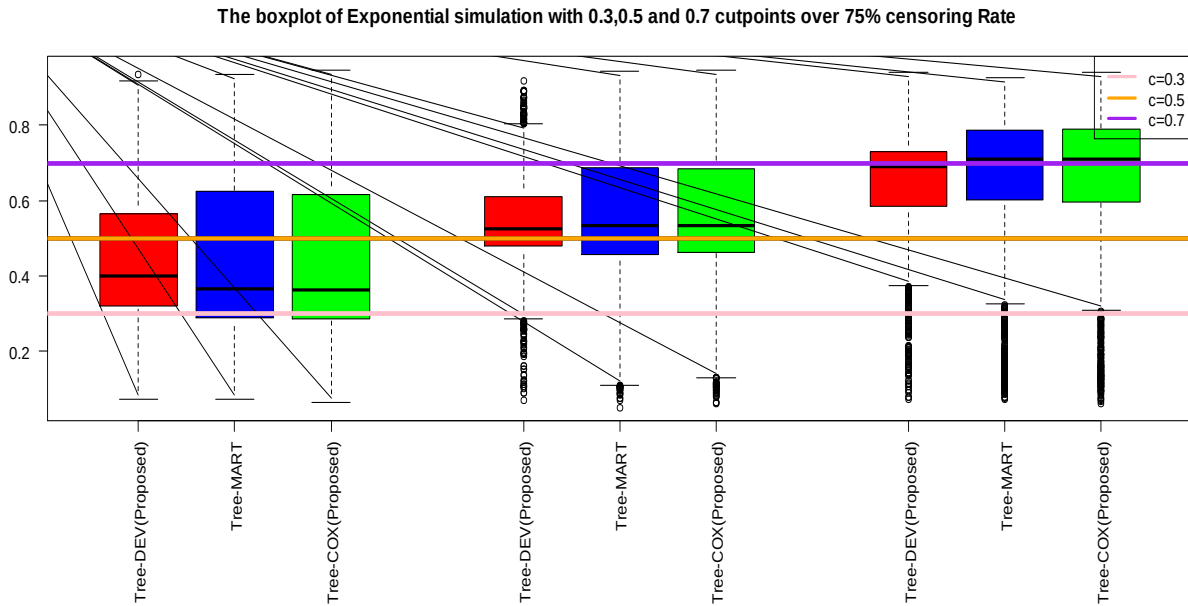


Fig. 4.4: Boxplot of exponential distribution

ii. **Results of Log-normal Distribution over the various censoring rates (25%, 50% & 75%)**

Simulation studies were conducted on models in Table 3.4, which was adopted to assess the performance of the proposed model in terms of cut point determination, the simulation was formulated based on log-normal distribution.

Similar to the exponential simulation above, the performances of the proposed models were as well examined based on mean and median, and their results were compared to the existing methodologies. Actually, the proposed model estimated values show in Table 4.4, 4.5 and 4.6 give a value that is much closer to the cutpoint selected than the existing techniques established by log-normal distribution, especially when the censoring rate is 50% and cutpoint chosen is 0.5 (i.e. moderate censoring). For instance, in Table 4.5 the mean and median of MRSS are 0.5467 and 0.5185 while DRSS(Proposed) are 0.4939 and 0.4900. Nevertheless, the existing method seems to behave better under log-normal distribution when the selected cutpoint are 0.1 and 0.9 under low censoring, 0.9 under moderate censoring and 0.9 under high censoring. Apart from these mentioned point, the proposed methods out shine the existing one.

Table 4.4: Low censoring rate 25%

|                 | Cutpoint c= 0.1 |               |               |               |               |
|-----------------|-----------------|---------------|---------------|---------------|---------------|
|                 | Mean            | Median        | SD            | RMSE          | MAE           |
| MRSS            | <b>0.3502</b>   | <b>0.2196</b> | 0.2694        | 0.3676        | <b>0.2515</b> |
| MRAV            | 0.5393          | 0.5550        | 0.2168        | 0.4824        | 0.4393        |
| DRSS(Proposed)  | 0.3521          | 0.2277        | 0.2656        | <b>0.3661</b> | 0.2529        |
| DRAV(Proposed)  | 0.4351          | 0.4151        | <b>0.2168</b> | 0.3991        | 0.3351        |
| CSRSS(Proposed) | 0.4795          | 0.4528        | 0.2835        | 0.4736        | 0.3809        |
| CSRAV(Proposed) | 0.4932          | 0.4883        | 0.2221        | 0.4516        | 0.3935        |
|                 | Cutpoint c= 0.5 |               |               |               |               |
| MRSS            | 0.5467          | 0.5185        | 0.1344        | 0.1422        | 0.0928        |
| MRAV            | 0.5809          | 0.5499        | 0.1094        | 0.1360        | 0.0955        |
| DRSS(Proposed)  | 0.5526          | 0.5207        | 0.1293        | 0.1396        | 0.0900        |
| DRAV(Proposed)  | 0.5377          | 0.5189        | <b>0.0835</b> | <b>0.0916</b> | <b>0.0595</b> |
| CSRSS(Proposed) | 0.4878          | 0.4822        | 0.2675        | 0.2677        | 0.2350        |
| CSRAV(Proposed) | <b>0.4939</b>   | <b>0.4900</b> | 0.2041        | 0.2042        | 0.1700        |
|                 | Cutpoint c= 0.9 |               |               |               |               |
| MRSS            | 0.7472          | 0.8677        | 0.2385        | 0.2832        | 0.1570        |
| MRAV            | <b>0.8463</b>   | <b>0.8810</b> | <b>0.1360</b> | <b>0.1462</b> | 0.7467        |
| DRSS(Proposed)  | 0.7669          | 0.8737        | 0.2245        | 0.2609        | <b>0.1376</b> |
| DRAV(Proposed)  | 0.7382          | 0.8350        | 0.2092        | 0.2644        | 0.6384        |
| CSRSS(Proposed) | 0.5092          | 0.5304        | 0.2710        | 0.4755        | 0.3918        |
| CSRAV(Proposed) | 0.5008          | 0.5075        | 0.2124        | 0.4521        | 0.4011        |

Table 4.5: Moderate censoring rate 50%

|                 | Cutpoint c= 0.1 |               |               |               |               |
|-----------------|-----------------|---------------|---------------|---------------|---------------|
|                 | Mean            | Median        | SD            | RMSE          | MAE           |
| MRSS            | 0.3168          | 0.1729        | 0.2639        | 0.3415        | 0.2182        |
| MRAV            | 0.4699          | 0.4628        | 0.2192        | 0.4246        | 0.3699        |
| DRSS(Proposed)  | <b>0.2814</b>   | <b>0.1568</b> | 0.2438        | <b>0.3038</b> | <b>0.1823</b> |
| DRAV(Proposed)  | 0.3793          | 0.3277        | 0.2192        | 0.3549        | 0.2794        |
| CSRSS(Proposed) | 0.4895          | 0.4714        | 0.2773        | 0.4781        | 0.3904        |
| CSRAV(Proposed) | 0.4981          | 0.4980        | <b>0.2126</b> | 0.4513        | 0.3984        |
|                 | Cutpoint c= 0.5 |               |               |               |               |
| MRSS            | 0.5409          | 0.5175        | 0.0809        | 0.0907        | 0.0525        |
| MRAV            | 0.5449          | 0.5247        | 0.0594        | 0.0745        | 0.0483        |
| DRSS(Proposed)  | 0.5421          | 0.5173        | 0.0778        | 0.0884        | 0.0501        |
| DRAV(Proposed)  | 0.5290          | <b>0.5150</b> | <b>0.0475</b> | <b>0.0556</b> | <b>0.0348</b> |
| CSRSS(Proposed) | 0.5259          | 0.5354        | 0.2458        | 0.2471        | 0.2084        |
| CSRAV(Proposed) | <b>0.5134</b>   | 0.5191        | 0.1847        | 0.1852        | 0.1468        |
|                 | Cutpoint c= 0.9 |               |               |               |               |
| MRSS            | 0.8718          | 0.8939        | 0.0944        | 0.0985        | 0.0342        |
| MRAV            | <b>0.8897</b>   | <b>0.8960</b> | <b>0.0245</b> | <b>0.0266</b> | 0.7897        |
| DRSS(Proposed)  | 0.8722          | 0.8942        | 0.0974        | 0.1013        | <b>0.0340</b> |
| DRAV(Proposed)  | 0.8526          | 0.8910        | 0.1181        | 0.1272        | 0.7526        |
| CSRSS(Proposed) | 0.5664          | 0.6119        | 0.2847        | 0.4386        | 0.3363        |
| CSRAV(Proposed) | 0.5345          | 0.5452        | 0.2215        | 0.4273        | 0.4347        |

Table 4.6: High censoring rate 75%

|                 | Cutpoint $c = 0.1$  |               |               |               |               |
|-----------------|---------------------|---------------|---------------|---------------|---------------|
|                 | Mean                | Median        | SD            | RMSE          | MAE           |
| MRSS            | <b>0.2815</b>       | <b>0.1562</b> | 0.2455        | 0.3053        | <b>0.1831</b> |
| MRAV            | 0.5644              | 0.5904        | <b>0.2135</b> | 0.5048        | 0.4644        |
| DRSS(Proposed)  | 0.2839              | 0.1628        | 0.2419        | <b>0.3038</b> | 0.1848        |
| DRAV(Proposed)  | 0.3958              | 0.3603        | <b>0.2135</b> | 0.3648        | 0.2959        |
| CSRSS(Proposed) | 0.4636              | 0.4318        | 0.2893        | 0.4646        | 0.3652        |
| CSRAV(Proposed) | 0.4777              | 0.4643        | 0.2268        | 0.4405        | 0.3780        |
|                 | Cut point $c = 0.5$ |               |               |               |               |
| MRSS            | 0.5431              | 0.5167        | 0.0841        | 0.0944        | 0.0544        |
| MRAV            | 0.5669              | 0.5422        | 0.0765        | 0.1016        | 0.0703        |
| DRSS(Proposed)  | 0.5497              | <b>0.5193</b> | 0.0864        | 0.0997        | 0.0572        |
| DRAV(Proposed)  | 0.5366              | 0.5198        | <b>0.0526</b> | <b>0.0641</b> | <b>0.0415</b> |
| CSRSS(Proposed) | 0.4755              | 0.4685        | 0.2585        | 0.2595        | 0.2248        |
| CSRAV(Proposed) | <b>0.4840</b>       | 0.4801        | 0.1936        | 0.1942        | 0.1590        |
|                 | Cutpoint $c = 0.9$  |               |               |               |               |
| MRSS            | 0.8451              | 0.8907        | 0.1459        | 0.1558        | 0.0614        |
| MRAV            | <b>0.8941</b>       | <b>0.8993</b> | <b>0.0221</b> | <b>0.0229</b> | 0.7941        |
| DRSS(Proposed)  | 0.8526              | 0.8920        | 0.1360        | 0.1440        | <b>0.0542</b> |
| DRAV(Proposed)  | 0.8112              | 0.8842        | 0.1599        | 0.1829        | 0.7113        |
| CSRSS(Proposed) | 0.5011              | 0.5158        | 0.2691        | 0.4811        | 0.3994        |
| CSRAV(Proposed) | 0.5013              | 0.5098        | 0.2104        | 0.4508        | 0.4014        |

The simulation results under log-normal distribution were further presented using: Root Mean Square Errors (RMSE) and Mean Absolute Errors (MAE) as shown in Figs. 4.5, 4.6 and 4.7. The various values of selected cut points varying from 0.1 to 0.9 were plotted against RMSE and MAE. This was done to evaluate the performances and the efficacy of the proposed models based on different levels of censoring rates. The figures showcase the efficiency of the proposed model over the existing ones since the blue line that represents the proposed model was found below the existing ones under low censoring rate irrespective of the impurity function. On the other hand, the existing model competes with the proposed model when the censoring rates were moderate and high. But, when the censoring rate was low the proposed model outperforms the existing techniques sum of square under RMSE and MAE.

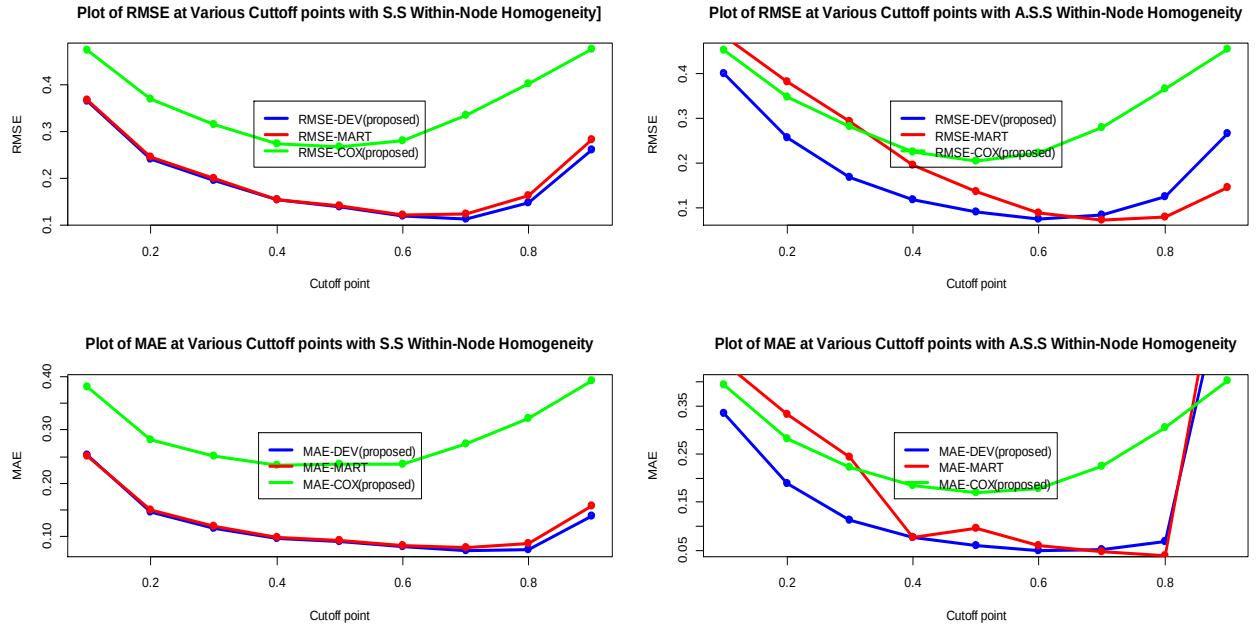


Fig. 4.5: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 25% censoring rate (Lognormal Distribution)

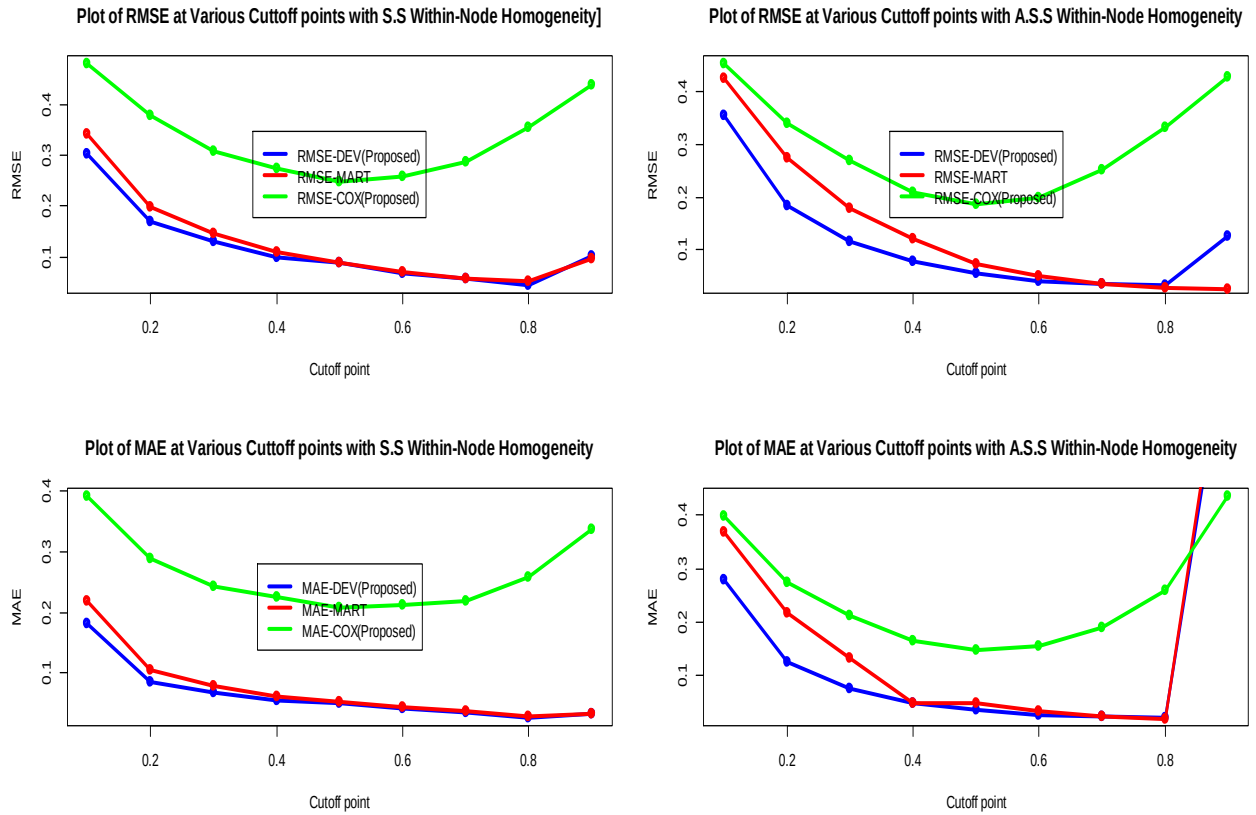


Fig. 4.6: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 50% censoring rate (Lognormal Distribution)

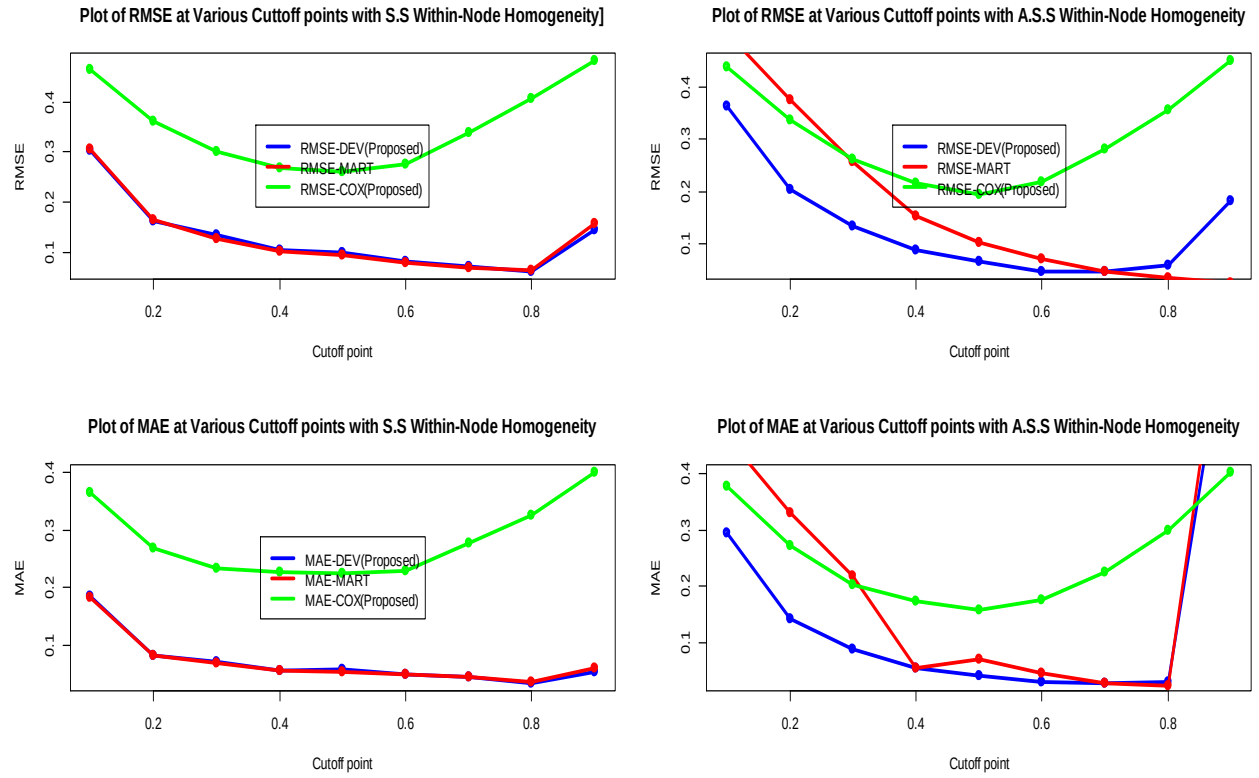


Fig. 4.7: The plot of the various cutpoints against the Root Mean Square Errors and Mean Absolute Errors at 75% censoring rate (Lognormal Distribution)

#### 4.1.2 Data Structure Performance results on survival based-tree via within-node homogeneity

Table 4.7 & 4.8 shows the result from examining the structure of our constructed trees based on exponential and log-normal distributions. The results measure the amount of variation encodes by the proposed methodology and existing one, their various results can be interpreted as follows: The bold-valued in these tables represent the ideal number of nodes, for multiplicative hazard model, three terminal nodes are expected and in additive hazard model 4 terminal nodes are of interest (Juanjuan; 2009 and Fiona, 2008). The average Mean Absolute Error was considered to check the performance of our grown tree. Results from these simulations revealed that the proposed models fit the data structure better than the existing methods based on the rate at which they detect the terminal nodes. Additionally, the Mean Absolute Error of the proposed models both on

exponential and log-normal distributions, were comparable to those presented in Fiona (2008). It is not surprising that the performance deteriorates a little with heavier censoring. But the location of the true cutoff point does not seem to have much effect.

Table 4.7: Examining tree structure based on 1 event of interest, through exponential distribution. N=500

| +                                                                                                   | Percent terminal nodes in final tree |           |           |   |   |    |          | Average       |               |
|-----------------------------------------------------------------------------------------------------|--------------------------------------|-----------|-----------|---|---|----|----------|---------------|---------------|
|                                                                                                     | 1                                    | 2         | 3         | 4 | 5 | 6  | $\geq 7$ | MAE1          | MAE2          |
| <b>Model 1: Exponential <math>\lambda = 0.1 + 0.3I\{(Z_1 &gt; 0.5) \cap (Z_2 &gt; 0.5)\}</math></b> |                                      |           |           |   |   |    |          |               |               |
| MRSS                                                                                                | 13                                   | 3         | 4         | 4 | 3 | 13 | 3        | 1.5336        | 1.5098        |
| MRAV                                                                                                | 60                                   | 14        | 13        | 3 | 3 | 4  | 3        | 1.5340        | 1.5190        |
| DRSS(Proposed)                                                                                      | 59                                   | 13        | 13        | 4 | 4 | 4  | 3        | 1.5310        | <b>1.5095</b> |
| DRAV(Proposed)                                                                                      | 60                                   | 15        | <b>14</b> | 3 | 3 | 3  | 2        | 1.5318        | 1.5160        |
| CSRSS(Proposed)                                                                                     | 64                                   | 7         | 4         | 7 | 5 | 6  | 7        | 1.5273        | 1.5185        |
| CSRAV(Proposed)                                                                                     | 86                                   | 4         | 1         | 1 | 4 | 4  | 0        | <b>1.5246</b> | 1.5172        |
| <b>Model 2: Exponential <math>\lambda = 0.05 + 0.2I(Z_1 &gt; 0.5) + (Z_2 &gt; 0.5)</math></b>       |                                      |           |           |   |   |    |          |               |               |
| MRSS                                                                                                | 60                                   | 12        | 11        | 5 | 4 | 4  | 4        | 1.5143        | 1.5265        |
| MRAV                                                                                                | 79                                   | 6         | 5         | 3 | 2 | 3  | 2        | 1.5134        | 1.5224        |
| DRSS(Proposed)                                                                                      | 60                                   | <b>13</b> | 10        | 5 | 4 | 4  | 4        | <b>1.5025</b> | 1.5268        |
| DRAV(Proposed)                                                                                      | 70                                   | 10        | 8         | 4 | 2 | 3  | 3        | 1.5144        | 1.5235        |
| CSRSS(Proposed)                                                                                     | 66                                   | 6         | 6         | 5 | 6 | 6  | 5        | 1.5131        | 1.5237        |
| CSRAV(Proposed)                                                                                     | 89                                   | 2         | 2         | 3 | 0 | 2  | 2        | 1.5112        | <b>1.5205</b> |

Table 4.8: Examining tree structure with one event-of-interest, through log-normal distribution. N=500

|                                                                                                        | Percent terminal nodes in final tree |    |           |          |     |     |          | Average       |               |
|--------------------------------------------------------------------------------------------------------|--------------------------------------|----|-----------|----------|-----|-----|----------|---------------|---------------|
|                                                                                                        | 1                                    | 2  | 3         | 4        | 5   | 6   | $\geq 7$ | MAE1          | MAE2          |
| <b>Model 3: Log normal <math>\mu = 2 + 1.5I\{(Z_1 &gt; 0.5) \cap (Z_2 &gt; 0)\}, \sigma = 1</math></b> |                                      |    |           |          |     |     |          |               |               |
| MSS                                                                                                    | 60                                   | 13 | 13        | 3        | 4   | 3   | 4        | 1.6826        | 2.0103        |
| MABS                                                                                                   | 60                                   | 14 | <b>14</b> | 3        | 3   | 2   | 4        | 1.6841        | 2.0121        |
| DSS(Proposed)                                                                                          | 60                                   | 13 | 13        | 3        | 4   | 3   | 4        | <b>1.6826</b> | <b>2.0075</b> |
| DABS(Proposed)                                                                                         | 59                                   | 14 | <b>14</b> | 3        | 4   | 3   | 3        | 1.6829        | 2.0102        |
| CSS(Proposed)                                                                                          | 62                                   | 13 | 7         | 5        | 4   | 5   | 4        | 1.7326        | 2.0208        |
| CABS(Proposed)                                                                                         | 78                                   | 15 | 5         | 0        | 0.5 | 0.5 | 1        | 1.7581        | 2.0127        |
| <b>Model 4: Log normal <math>\mu = 2 - 0.85I(Z_1 &gt; 0.5) + 0.85Z_2, \sigma = 1</math></b>            |                                      |    |           |          |     |     |          |               |               |
| MSS                                                                                                    | 59                                   | 13 | 13        | 3        | 4   | 5   | 3        | 1.7128        | 1.9956        |
| MABS                                                                                                   | 60                                   | 14 | 13        | 2        | 4   | 4   | 3        | 1.7123        | <b>1.9867</b> |
| DSS(Proposed)                                                                                          | 60                                   | 12 | 12        | 3        | 5   | 5   | 3        | 1.7123        | 2.0007        |
| DABS(Proposed)                                                                                         | 60                                   | 13 | 13        | 3        | 4   | 4   | 3        | <b>1.7097</b> | 1.9904        |
| CSS(Proposed)                                                                                          | 63                                   | 12 | 6         | <b>5</b> | 5   | 5   | 4        | 1.7596        | 1.9988        |
| CABS(Proposed)                                                                                         | 78                                   | 14 | 3         | 1        | 1   | 2   | 1        | 1.7624        | 1.9946        |

### 4.1.3 Application of survival based-tree via within-node homogeneity on Real-life Dataset

This section contains real-life data designed to complement the simulation results, and aimed at investigating the performance of the proposed methods and the existing methods as well as identifying variables that are important and associated to survival of a patient from any kind of disease biomarker. The two real-life data considered are: bone marrow transplant and Prostate cancer.

#### i. Bone marrow transplant Dataset

The bone marrow transplant study was introduced in the package “timereg” in R (see the attached CD for the raw and complete data). The data consist of two competing risks (relapse and death in remission) and three covariates, which include: Platelet, Tcell and Age, and they are all coded as follows:

Platelet: a numeric vector code. Platelet 1: more than  $100 \times 10^9$  per L, 0: less.

Tcell: a numeric vector. T-cell depleted BMT 1: yes, 0: no.

Age: a numeric vector code. Age of patient, scaled and centered  $((\text{age}-35)/15)$ .

The length of stay in remission and the causes were also coded below:

Cause: a numeric vector code. Survival status. 1: death in remission, 2: relapse, 0: censored.

Time: a numeric vector. Survival time.

We now apply the proposed tree method to develop prognostic classification rule using the bone marrow transplant. We first considered the risk of relapse (cause 1) and subsequently, the risk of death in remission (cause 2) using the proposed model as well as the existing model and yielded the results in Tables 4.9 and 4.10 and figs. 4.8 and 4.9.

Table 4.9: Results of bone marrow transplant through cause 1 (relapse)

| Models                                   | Deviance       | Number of terminal nodes | The first covariate to be split | Mean of Squared Residuals |
|------------------------------------------|----------------|--------------------------|---------------------------------|---------------------------|
| Proposed model 1<br>(Deviance Residual)  | 556.8822       | 4                        | Age                             | 1.2703                    |
| Proposed model 2<br>(Cox-Snell Residual) | <b>16.6498</b> | 7                        | Age                             | <b>0.0376</b>             |
| Existing Model<br>(Martingale Residual)  | 160.3592       | 4                        | Age                             | 0.3669                    |

In Table 4.9, the results of prognostic classification tree based on bone marrow transplant for various methods were presented when relapse (cause 1) was considered as an event of interest. The deviance results were used as a measure of model selection and models with the lowest Deviance values are generally preferred over other models. It was observed that the proposed model (Cox-Snell residual) via within-node homogeneity provides a deviance value that is lower than other existing methods. Interestingly, when the mean squared error was considered as a measure of model selection, it yielded a result that is similar to other criterion.

Moreover, the virtual display in figures 4.8, 4.10 and 4.12 revealed that the explanatory variable age plays a key role in constructed classification trees along with other covariates therein irrespective of the models. This suggests that the survival wood of patients with bone marrow transplant rest on the age of that patient while other covariates (factors) follow.

Another important statistical tool to be considered is called variable importance. The variable importance allocates uppermost values to factors with the utmost divergence between the original forecast performance and prediction performance after transformation. The resulting variable importance values have no meaning on an absolute scale, but their relative sizes can be useful for comparing across different predictor variables (for full details see Olshen 2010 and Breiman 2001). Results of figures. 4.8, 4.10 and 4.12 revealed that age is an important factor in bone

marrow transplant when relapse is of interest since it has longer bar than other factors such as platelets and tcells. Additionally, the left-hand sides of figs. 4.9, 4.11 and 4.13 represent the possible number of tree grown before the tree converges (stabilized). The conversion was observed to be at around 200 trees at was shown and this authenticates the 200 number trees grown in the simulation studies.

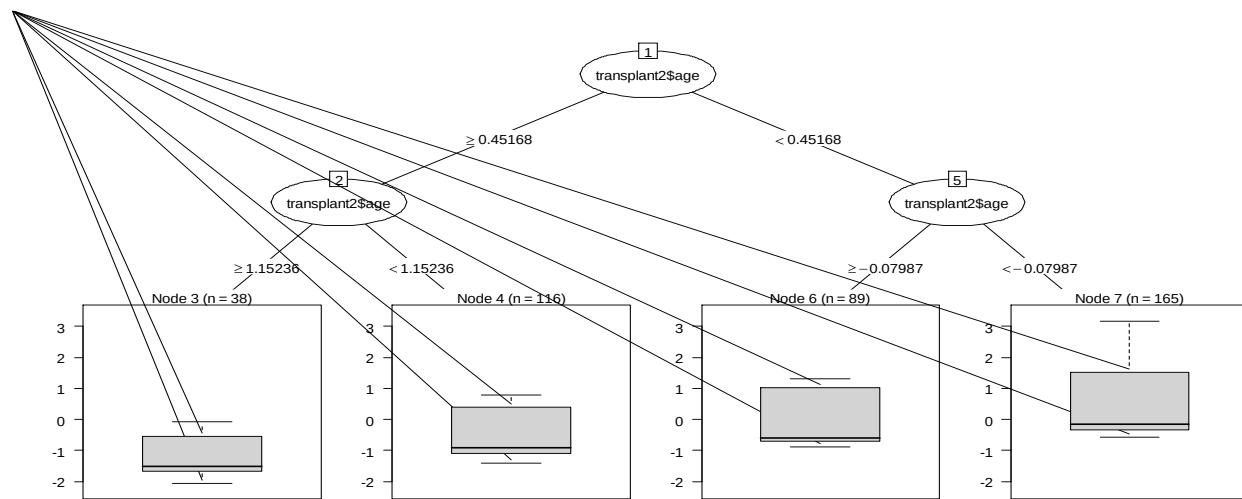


Fig. 4.8: Within node tree for data from bone marrow transplant through cause 1 (relapse) via martingale residual

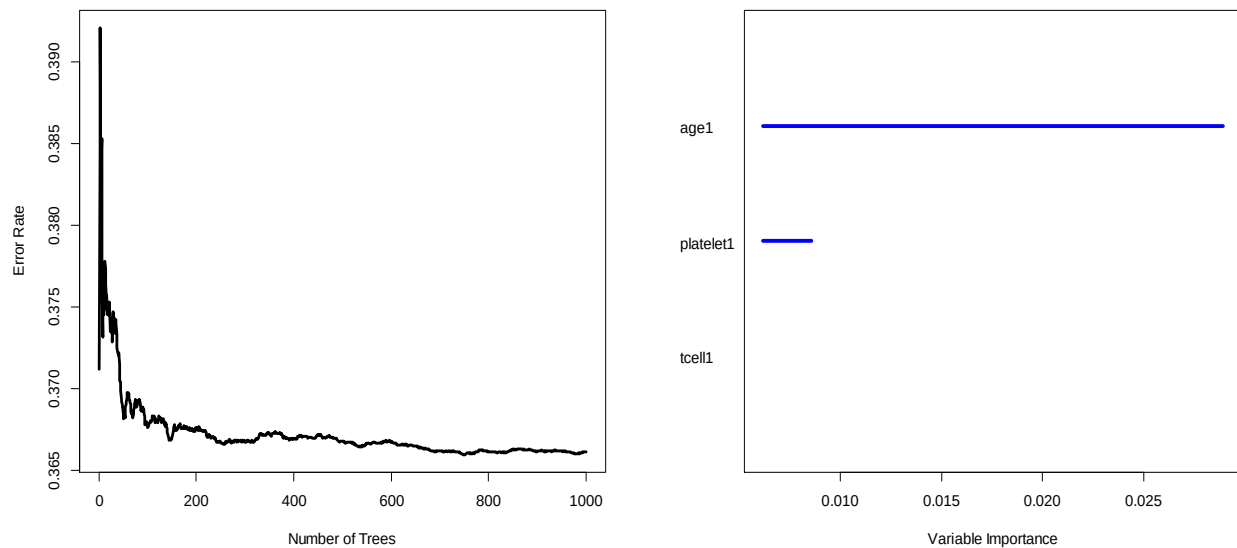


Fig. 4.9: Graphical representation of the variable importance in bone marrow transplant through cause 1 (relapse) via martingale residual

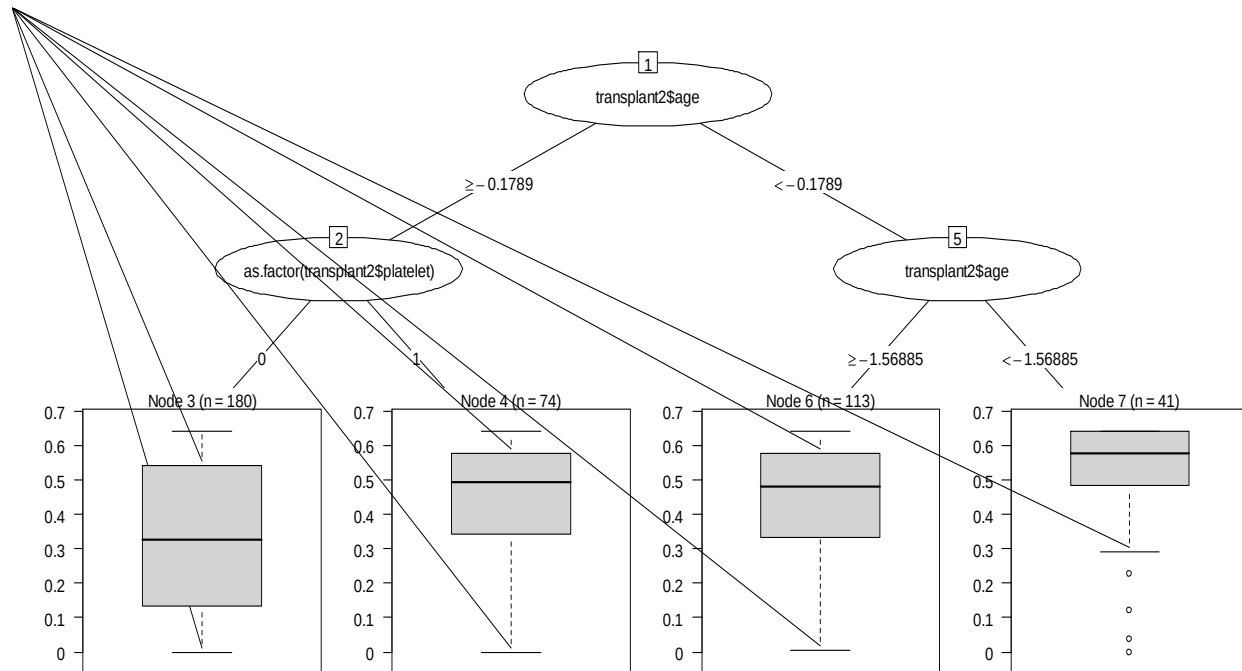


Fig. 4.10: Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual

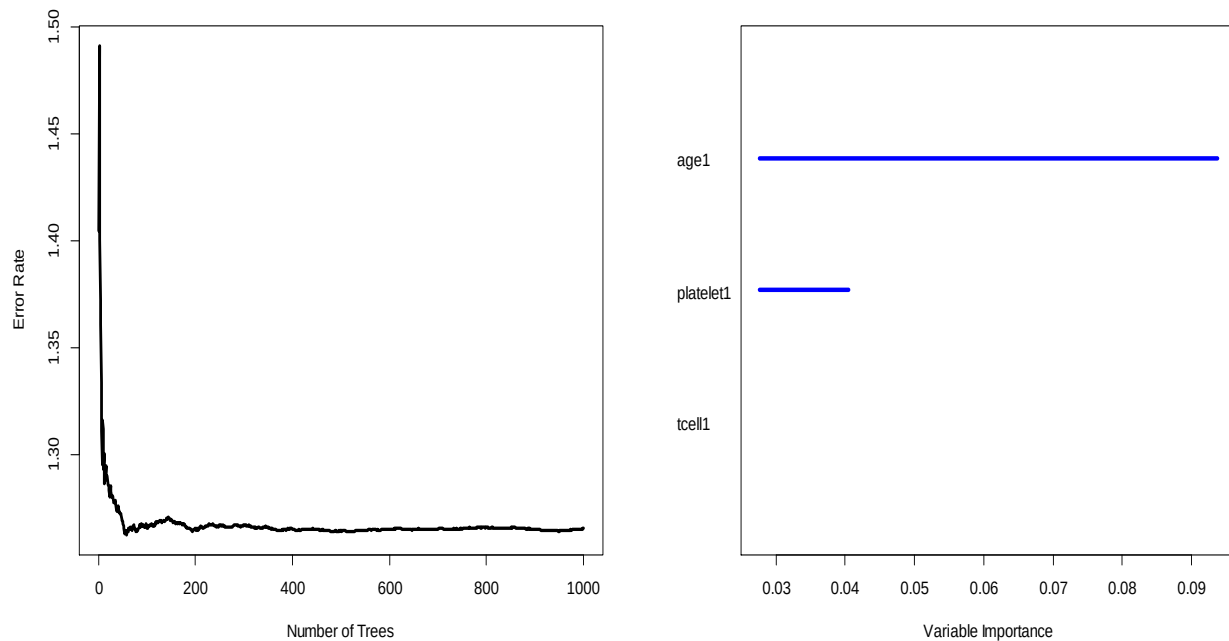


Fig. 4.11: Graphical representation of the variable importance in bone marrow transplant through cause 1 (relapse) via deviance residual.

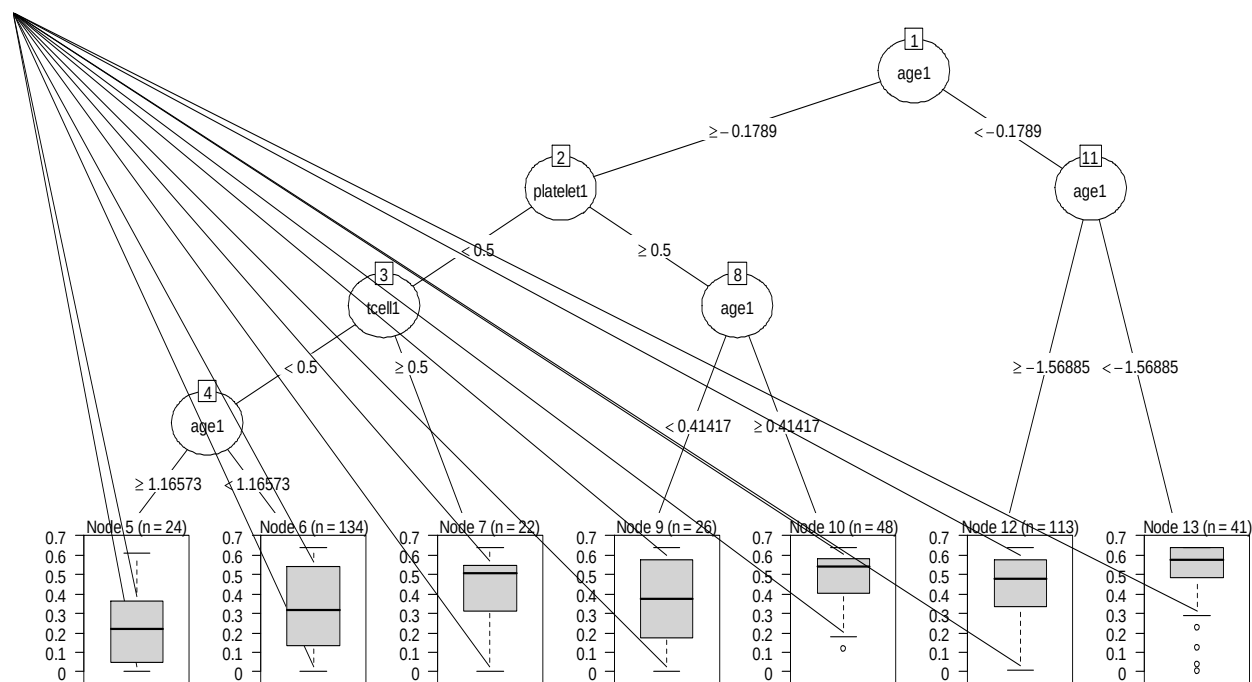


Fig. 4. 12: Within node tree for data from bone marrow transplant through cause 1 (relapse) via Cox-Snell residual

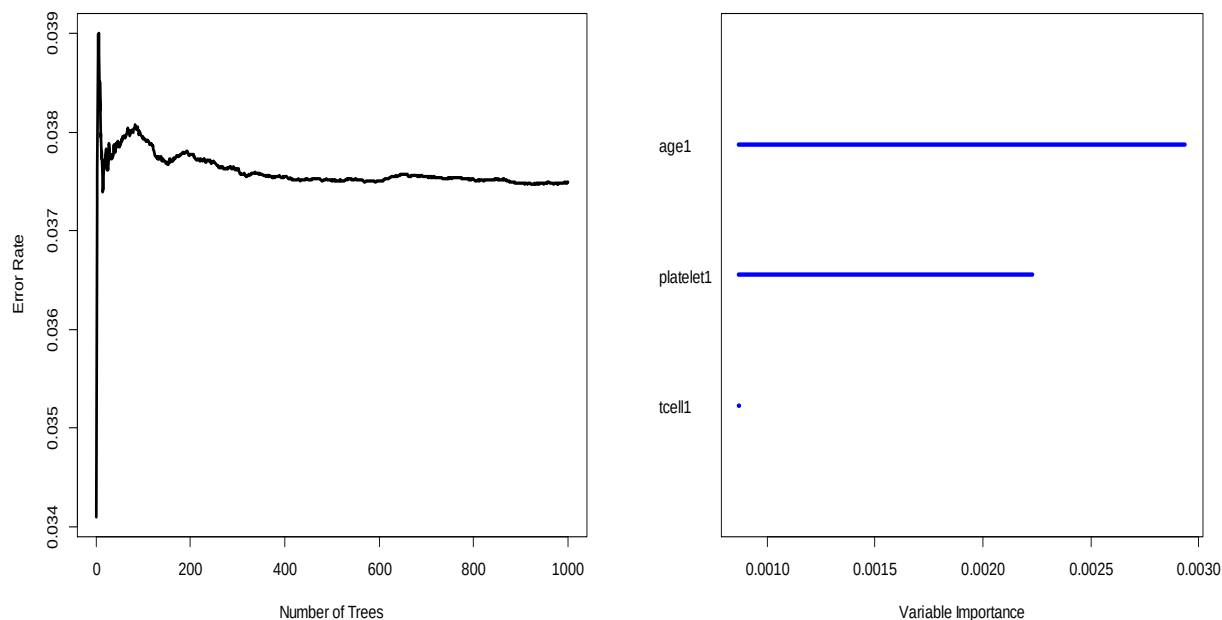


Fig. 4.13: Graphical representation of the variable importance in bone marrow transplant through cause 1 (relapse) via Cox-Snell residual

Table 4.10: Results of bone marrow transplant through cause 2 (death in remission)

| <b>Models</b>                            | <b>Deviance</b> | <b>Number of terminal nodes</b> | <b>The first covariate to be split</b> | <b>Mean Squared Residuals</b> |
|------------------------------------------|-----------------|---------------------------------|----------------------------------------|-------------------------------|
| Proposed model 1<br>(Deviance Residual)  | 372.4276        | 3                               | Age                                    | 0.9120                        |
| Proposed model 2<br>(Cox-Snell Residual) | <b>49.5625</b>  | 3                               | Age                                    | <b>0.1175</b>                 |
| Existing Model<br>(Martingale Residual)  | 86.5535         | 4                               | Age                                    | 0.2120                        |

The second competing risk in the bone marrow transplant is death in remission; the prognostic classification tree results were shown in table 4.10. Equivalently, deviance and mean squared residuals were used to measure the performances of the proposed models and the existing model. Conversely, it was observed that the proposed model 2 (Cox-Snell residual) performed better than other empirical methods based on the two selected measurements mentioned earlier.

Besides, the tree displayed in figure 4.14, 4.16 and 4.18 revealed that the covariate age plays a key role in constructing prognostic classification trees since it is the first covariate to be splitted in the tree when remission is the variable of interest in bone marrow transplant while other explanatory variables follow. These results are similar to the results generated when relapse is of interest.

Finally, figures 4.15, 4.17 and 4.18 show how the dependent variables (age, platelet and Tcells) are important to outcome of death in remission based on a certain period of time. Consistently, age was found to be the most important covariate in determining the outcome of death in remission in bone marrow transplant. This is also similar to results of relapse when variable importance was considered.

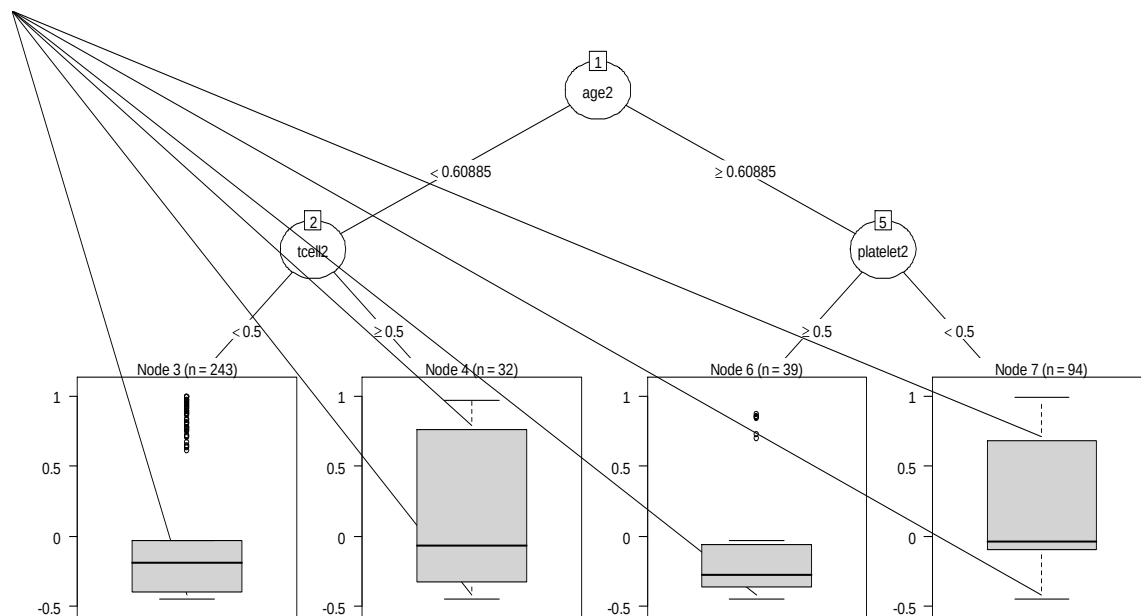


Fig. 4.14: Within node tree for data from bone marrow transplant through cause 2 (death in remission) via martingale residual

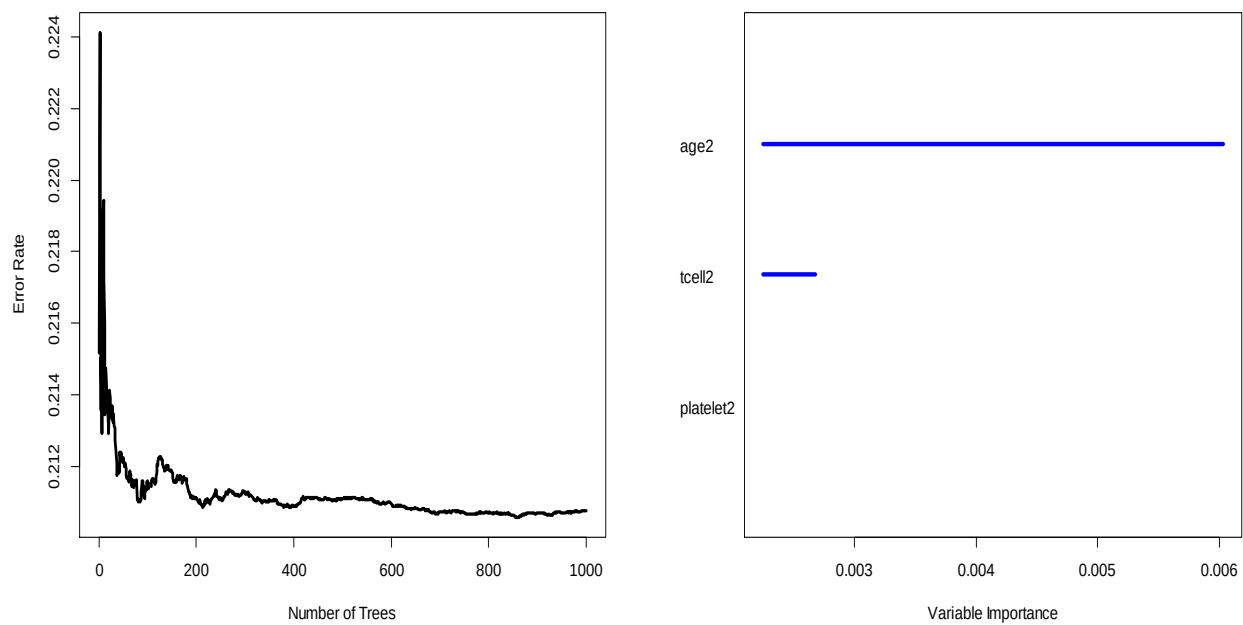


Fig. 4.15: Graphical representation of the variable importance in bone marrow transplant through cause 1 (relapse) via martingale residual

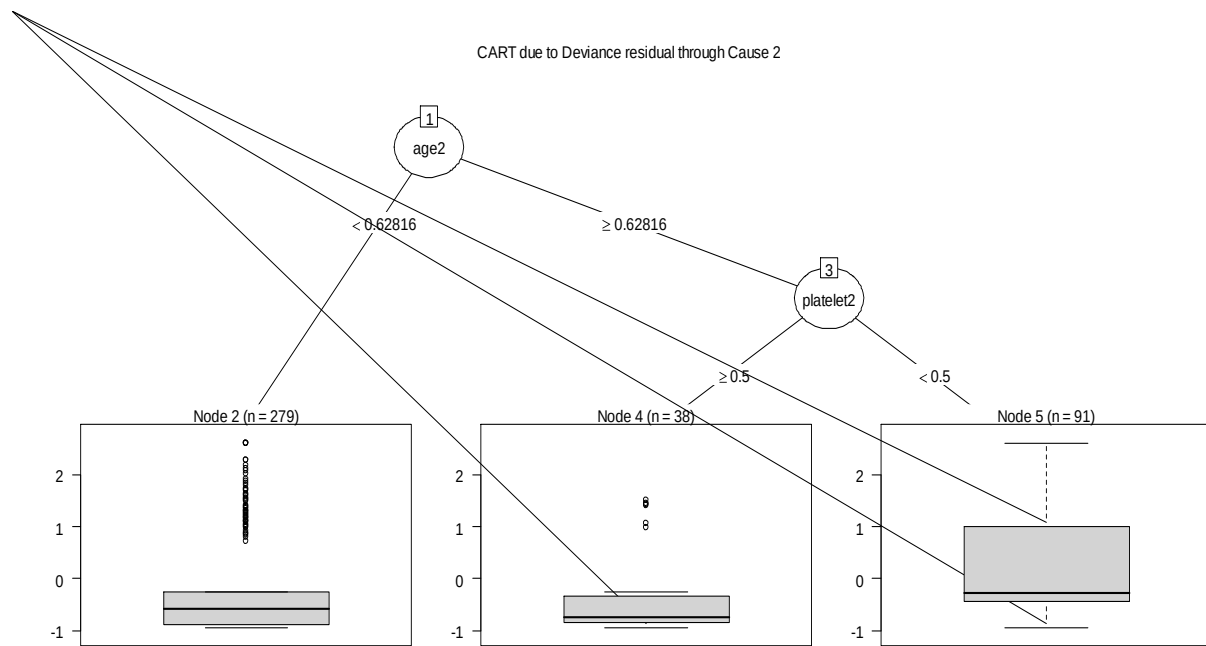


Fig. 4.16: Within node tree for data from bone marrow transplant through cause 1 (relapse) via deviance residual

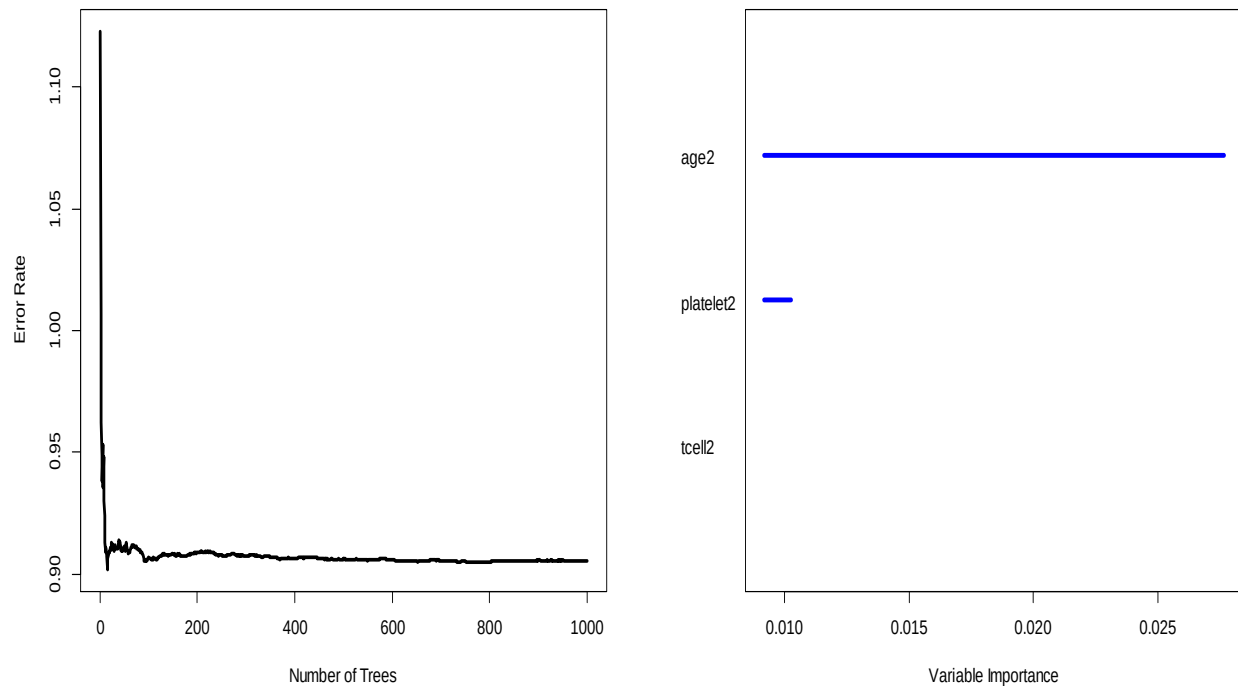


Fig. 4.17: Graphical representation of the variable importance in bone marrow transplant through cause 2 (death in remission) via martingale residual.

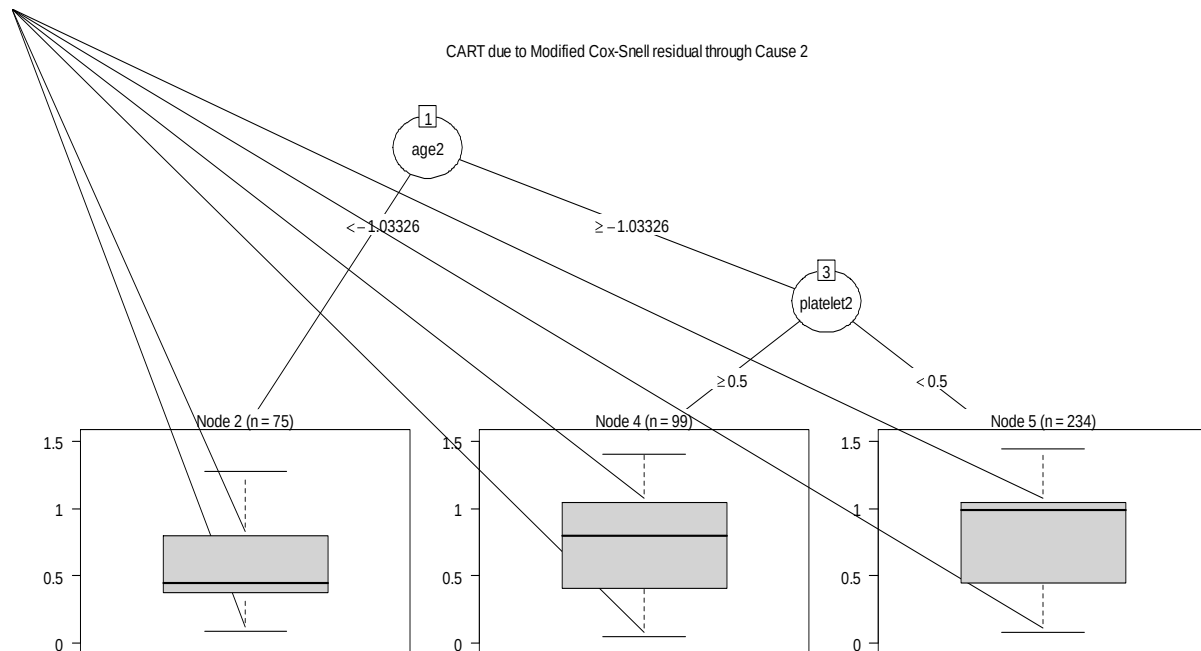


Fig. 4.18: Within node tree for data from bone marrow transplant through cause 2 (death in remission) via Cox-Snell residual

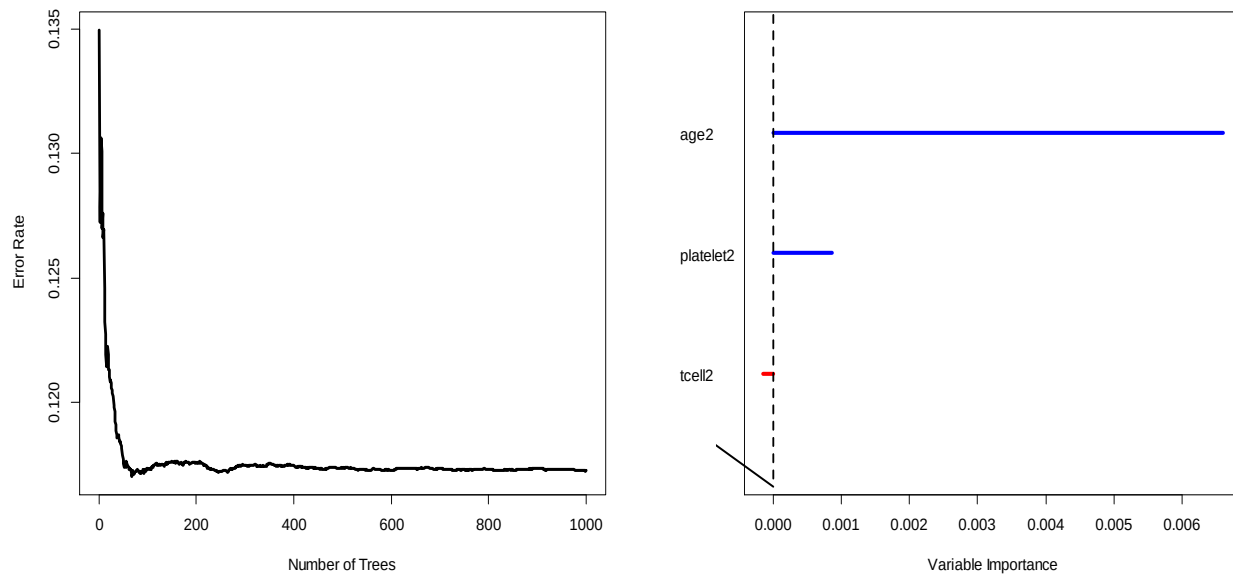


Fig. 4.19: Graphical representation of the variable important in bone marrow transplant through cause 2 (death in remission) via Cox-Snell residual

## ii. Prostate Cancer Data

The data contain double-blinded randomized clinical trial, conducted in order to compare two treatments for patients with prostate cancer. The two treatments were placebo and 1.0mg diethylstilbestrol (DES). Information on 236 patients were obtained of which 119 of them were given placebo, the remaining 117 patients were on DES. Apart from treatment, other prognostic factors were also reported in the dataset.

The main aims of this trial were to determine how survival time of patients was affected by the treatment and any other prognostic factor(s). Also, possible toxic effect of DES particularly in terms of mortality from heart disease was to be examined. However, endpoint considered was death from prostatic cancer and death from any other causes. The following covariates were used

for this research, others were discarded due to the fact that they are categorical variables: Age in years, Weight index (WTINDEX), and Serum Haemoglobin gm/100ml (SERUMHG). The Follow-up Time is in Months and the Survival Status was recorded by treating death from prostatic cancer as the event of interest (cause 1), death from Heart or Vascular Disease and other causes as competing risks (cause 2) and discharge alive as true censoring (event 0). The Descriptive statistics of the data were shown in Table 4.11 while the various results of prognostic classification tree based on cutpoint were presented in Table 4.12.

Table 4.11: The Summary Statistics of Age, Weight index (WTINDEX) and Serum Haemoglobin gm/100ml (SERUMHG) in prostate cancer.

| <b>Heart or Vascular Disease (event 1)</b> |       |        |         |         |
|--------------------------------------------|-------|--------|---------|---------|
| Variable                                   | Mean  | Median | Minimum | Maximum |
| Age                                        | 70.37 | 73.00  | 50.00   | 81.00   |
| WTINDEX                                    | 99.33 | 99.00  | 71.00   | 135.00  |
| SERUMHG                                    | 12.98 | 13.45  | 7.00    | 16.40   |
| <b>Others causes (event 2)</b>             |       |        |         |         |
| Age                                        | 73.15 | 74.00  | 51.00   | 84.00   |
| WTINDEX                                    | 98.28 | 96.50  | 72.00   | 152.00  |
| SERUMHG                                    | 13.42 | 13.60  | 5.90    | 18.20   |

Table 4.12: Cutpoint selected based on the terminal nodes using measure of impurities (event 1).

| <b>Heart or Vascular Disease (event 1)</b> |          |          |          |          |          |          |
|--------------------------------------------|----------|----------|----------|----------|----------|----------|
| Measures of Impurities                     | Age      |          | WTINDEX  |          | SERUMHG  |          |
|                                            | Cutpoint | Deviance | Cutpoint | Deviance | Cutpoint | Deviance |

|                                |               |                  |               |                  |          |          |
|--------------------------------|---------------|------------------|---------------|------------------|----------|----------|
| Proposed model (Deviance)      | 68.5          | 23.88400         | 99.5          | 30.73562         | 11.85 22 | 35.33052 |
| Proposed model (Cox-snell)     | <b>75.5**</b> | <b>0.9452955</b> | <b>92.5**</b> | <b>1.2398660</b> | ***      | ***      |
| Existing Model (Martingale)    | 68.5          | 6.4395           | 99.5          | 7.6823           | 11.85    | 7.1945   |
| <b>Others causes (event 2)</b> |               |                  |               |                  |          |          |
|                                | Age           |                  | WTINDEX       |                  | SERUMHG  |          |
|                                | Cutpoint      | Deviance         | Cutpoint      | Deviance         | Cutpoint | Deviance |
| Proposed model (Deviance)      | 67.5 43       | 34.47947         | 104.5         | 188.67250        | ***      | ***      |
| Proposed model (Cox-snell)     | 76.5          | 12.209020        | ***           | ***              | ***      | ***      |
| Existing Model (Martingale)    | 75.5          | 17.861760        | 104.5         | 24.024260        | ***      | ***      |

\*\* represents the best model based on the deviance and \*\*\* means SERUMHG or WTINDEX does not have

much effect on Heart or Vascular Disease and Other causes.

In general, there are two distinct methods of determining the cut-points and these fall into two categories. The first category is the use of data-oriented approach which involves the use of continuous covariates based on quantiles such as median or upper quartile. The second category is the use of outcome-oriented method, which provides a value of a cutpoint based on the most significant relation to outcome through the use of statistical tools such as the proposed prognostic classification tree. For full detail see Mandrekar et al. (2006).

Table 4.11 provides the data-oriented approach for this work based on the median of all the continuous variables: Age, Weight index (WTINDEX) and Serum Haemoglobin gm/100ml (SERUMHG) of death due to Heart or Vascular Disease (event 1) and Other causes (event 2) such as: Cerebrovascular accident, Thrombosis and Respiratory Disorder were 73 years, 99 and 13.45 gm/100ml for event 1 and 74 years, 96.50 and 13.60 gm/100ml for event 2. This can be used as a cutpoint for any patient treated with diethylstilbestrol (DES).

Table 4.12 presented the results of outcome-orientation based on the proposed prognostic classification tree. We considered the information about Heart or Vascular Disease and Other

causes, then, the results revealed that: The Proposed model (Cox-Snell) performed better than any other methods irrespective of the events. Moreover, the age and Weight index seems to be very important in prostate cancer with cutpoint of 76 years and 92.5 weighted indexes.

Subsequently, we considered the smoothed spline plot using the proposed and existing prognostic classification trees as the outcome-oriented approach as described earlier to determine the behavior of patient's age only.

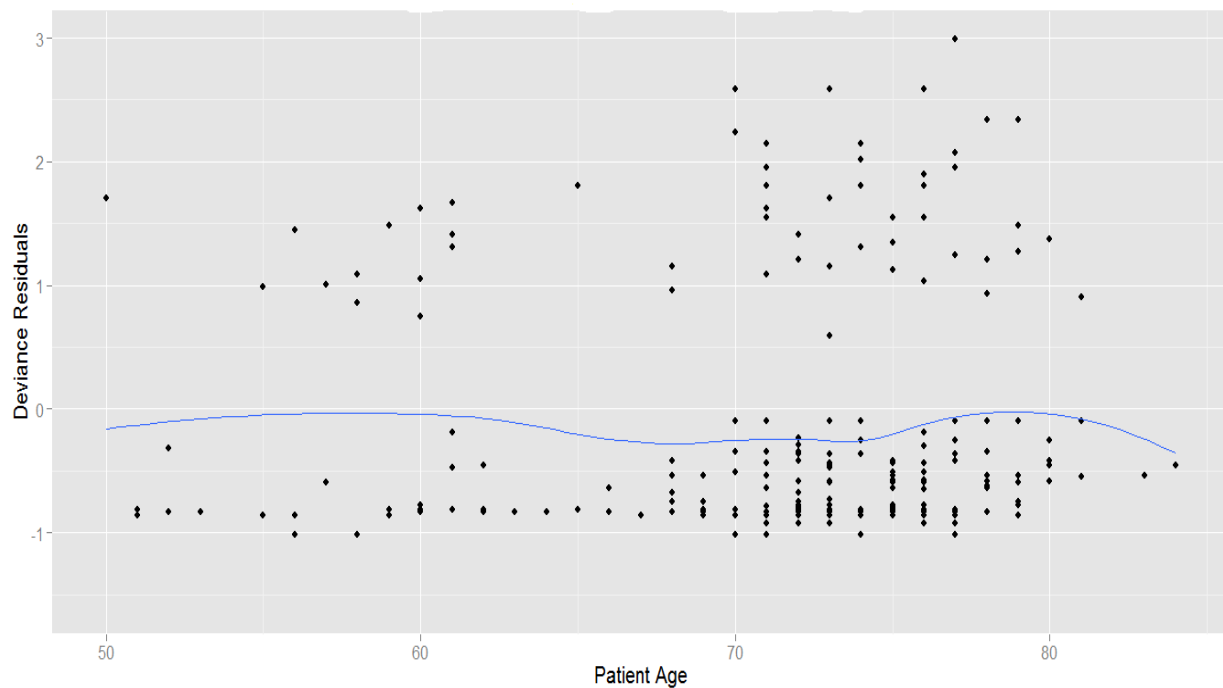


Fig. 4.20: Plot of Martingale residual against the patient Age using existing prognostics classification tree

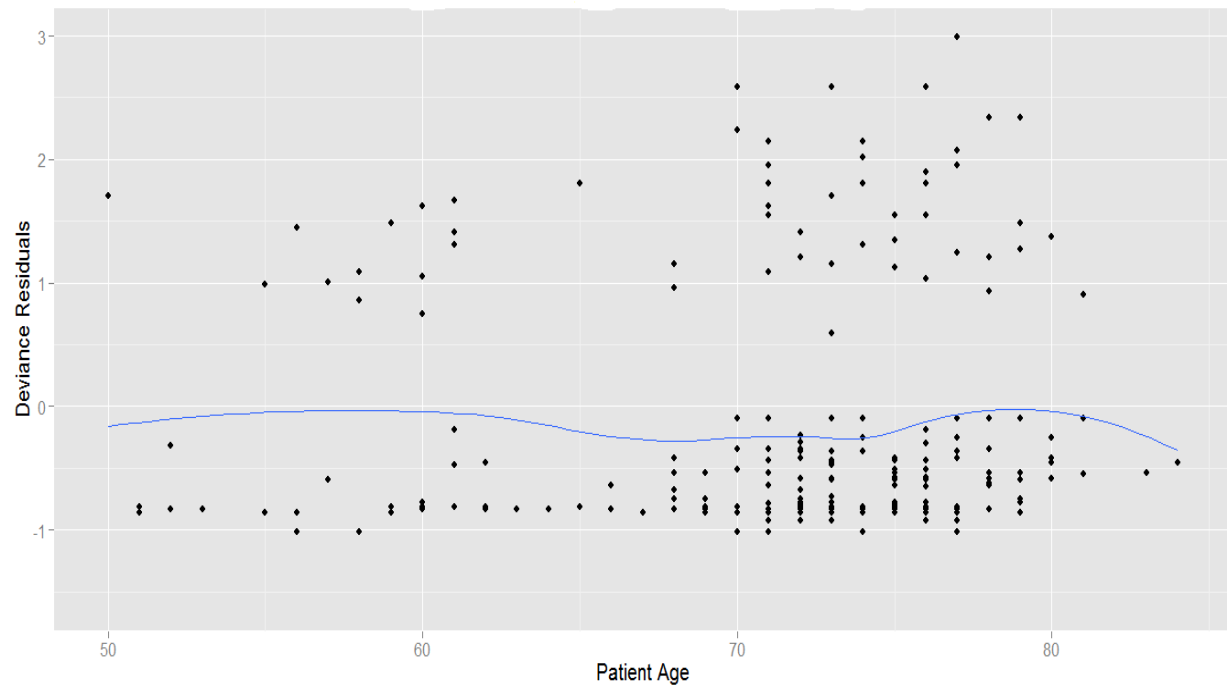


Fig. 4.21: Plot of Deviance residual against the patient Age using proposed prognostics classification tree

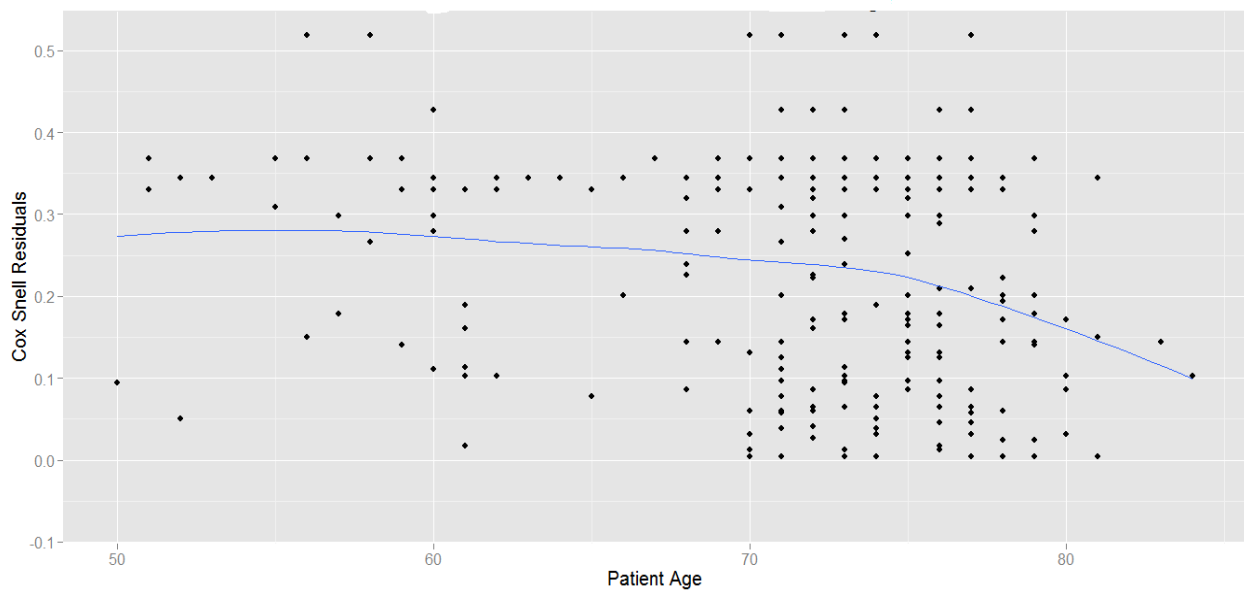


Fig. 4.22: Plot of Cox-Snell residual against the patient Age using proposed prognostics classification tree

The virtual presentation in fig. 4.20, 4.21, 4.22, illustrate the functional relationship that exists between the covariate (patient age) and outcome of interest (residual of the survival time). The results reveal the non-linear association that exists in the model and thus, treating patient age as a

linear relationship with the time and outcome in prostate cancer is inappropriate. In this regard, the classical approach will not be suitable for the analysis, but the proposed prognostic classification tree has the ability to handle linear and non-linear relationships that exist. These results were similar to Mandrekar et al. (2006).

## 4.2 Survival Tree-Based via Between-Node Heterogeneity (Task 2) Numerical Results

The simulation scheme of this task has been fully explained in chapter three. This study considers two survival distributions: exponential and log-normal distributions, their various numerical values and their explanations were given in subsequent subsections. The main method of assessment is mean and median, and their closeness to the cutpoint indicate a good performance of the various adopted splitting rules.

### i. Results of Exponential Distribution on Between-Node Heterogeneity (Task 2)

The various basic terms used in this study have been discussed in chapter one, but for the purpose of this subsection, here are the meanings of some of the abbreviations used in table 4.13:  $T^{NS}$  represent Neyman smooth statistic test,  $T^{KS}$  represents Kolmogorov Smirnov test,  $T^{Pepe}$  represents Pepe and Moris test,  $T^G$  represents Gray test,  $T^{FG}$  represents Fine and Gray test,  $T^{CPH}$  represents Cox-Proportional Hazard model and  $T^{LR}$  represents Log-rank test. The various mean and median of the test statistics were ranked based on their closeness to the chosen mean and median, at the same time they were ranked appropriately. The closest one was ranked first, and next to it was ranked second and third respectively, this was done differently according to the chosen sample sizes. It can be deduced from the results in table 4.13 that; Neyman smooth test was found to be more efficient in detecting the cutpoint of a data than any other methods including the existing

methods. However, the Cox-Proportional Hazard model and log-rank were found to perform least since they both have least ranks as shown in the table 4.13. The second method that performs relatively equally as Neyman smooth is pepe integral and next to it is Kolmogorov tests.

Table 4.13: Simulation with Exponential Distribution over 50% censoring rate and 0.5 cut-points

| Sample size | Performance Measures | Two-sample Tests of the Equality of Two Cumulative Incidence Functions |                             |                       |                          |                      |                      |                      |
|-------------|----------------------|------------------------------------------------------------------------|-----------------------------|-----------------------|--------------------------|----------------------|----------------------|----------------------|
|             |                      | $T^{NS*}$                                                              | $T^{KS*}$                   | $T^{Ppe*}$            | $T^{G(existing)}$        | $T^{FG*}$            | $T^{CPH*}$           | $T^{LR}$             |
| 20          | Mean                 | 0.5365(2)<br>[0.037]                                                   | <b>0.4834(1)</b><br>[0.017] | 0.4722(3)<br>[0.027]  | 0.5891(7)<br>[0.089]     | 0.5842(4)<br>[0.084] | 0.5493(5)<br>[0.049] | 0.5537(6)<br>[0.054] |
|             | Median               | 0.5480(3)<br>[0.048]                                                   | 0.4548(2)<br>[0.045]        | 0.4296(4)<br>[0.070]  | 0.5445(2)<br>[0.045]     | 0.5706(5)<br>[0.071] | 0.7799(6)<br>[0.280] | 0.5225(1)<br>[0.023] |
| 50          | Mean                 | 0.5164(1)<br>[0.015]                                                   | 0.5731(6)<br>[0.073]        | 0.5269(2)<br>[0.027]  | 0.5388(4)<br>[0.039]     | 0.5376(3)<br>[0.038] | 0.5707(5)<br>[0.071] | 0.5390(4)<br>[0.039] |
|             | Median               | 0.5105(1)<br>[0.011]                                                   | 0.6340(6)<br>[0.134]        | 0.6151(5)<br>[0.115]  | 0.5151(2)<br>[0.015]     | 0.5235(4)<br>[0.024] | 0.7799(7)<br>[0.280] | 0.5165(3)<br>[0.017] |
| 80          | Mean                 | 0.5147(1)<br>[0.015]                                                   | 0.5411(4)<br>[0.041]        | 0.5161(2)<br>[0.016]  | 0.5409(4)<br>[0.041]     | 0.5371(3)<br>[0.037] | 0.5493(5)<br>[0.049] | 0.5371(3)<br>[0.037] |
|             | Median               | 0.5050(1)<br>[0.005]                                                   | 0.5740(5)<br>[0.074]        | 0.5775(6)<br>[0.078]  | 0.5172(2)<br>[0.017]     | 0.5501(4)<br>[0.050] | 0.7539(7)<br>[0.254] | 0.5208(3)<br>[0.021] |
| 100         | Mean                 | 0.4838(2)<br>[0.016]                                                   | 0.5318(3)<br>[0.032]        | 0.5068 (1)<br>[0.007] | 0.5483(7)<br>[0.048]     | 0.5455(6)<br>[0.046] | 0.5452(5)<br>[0.045] | 0.5396(4)<br>[0.040] |
|             | Median               | 0.4725(3)<br>[0.028]                                                   | 0.5635(5)<br>[0.064]        | 0.5544(4)<br>[0.054]  | 0.5208<br>(1)<br>[0.021] | 0.5675(6)<br>[0.068] | 0.7444(7)<br>[0.244] | 0.5251(2)<br>[0.025] |
| 150         | Mean                 | 0.5004(1)<br>[0.000]                                                   | 0.5336(3)<br>[0.034]        | 0.5065(2)<br>[0.007]  | 0.5571(6)<br>[0.057]     | 0.5440(4)<br>[0.044] | 0.5537(5)<br>[0.054] | 0.5535(5)<br>[0.054] |
|             | Median               | 0.5027(1)<br>[0.003]                                                   | 0.5622(4)<br>[0.062]        | 0.5410(3)<br>[0.041]  | 0.5296(2)<br>[0.030]     | 0.5691(5)<br>[0.070] | 0.7539(6)<br>[0.254] | 0.5299(2)<br>[0.030] |
| 200         | Mean                 | 0.4952(1)<br>[0.005]                                                   | 0.5192(3)<br>[0.019]        | 0.4898(2)<br>[0.010]  | 0.5564(7)<br>[0.056]     | 0.5248(4)<br>[0.025] | 0.5522(6)<br>[0.052] | 0.5469(5)<br>[0.047] |
|             | Median               | 0.4986(1)<br>[0.001]                                                   | 0.5426(3)<br>[0.043]        | 0.4563(4)<br>[0.044]  | 0.5262(2)<br>[0.026]     | 0.5440(4)<br>[0.044] | 0.7444(5)<br>[0.244] | 0.5262(2)<br>[0.026] |

|      |        |                      |                       |                      |                      |                      |                       |                      |
|------|--------|----------------------|-----------------------|----------------------|----------------------|----------------------|-----------------------|----------------------|
| 500  | Mean   | 0.4857(3)<br>[0.014] | 0.4861(3)<br>[0.014]  | 0.4983(1)<br>[0.002] | 0.5510(4)<br>[0.051] | 0.5085(2)<br>[0.009] | 0.5420(5)<br>[0.042]  | 0.5433(6)<br>[0.043] |
|      | Median | 0.4789(1)<br>[0.021] | 0.4948(5)<br>[0.052]  | 0.4764(3)<br>[0.024] | 0.5250(4)<br>[0.025] | 0.5221(2)<br>[0.022] | 0.7192(6)<br>[0.2192] | 0.5220(2)<br>[0.022] |
| 1000 | Mean   | 0.4983(2)<br>[0.002] | 0.4877(4)<br>[0.012]  | 0.4931(3)<br>[0.007] | 0.5472(7)<br>[0.047] | 0.4997(1)<br>[0.000] | 0.5225(5)<br>[0.023]  | 0.5392(6)<br>[0.039] |
|      | Median | 0.5004(1)<br>[0.000] | 0.4978 (2)<br>[0.002] | 0.4727(6)<br>[0.027] | 0.5201(5)<br>[0.020] | 0.5050(3)<br>[0.005] | 0.6710(7)<br>[0.171]  | 0.5175(4)<br>[0.018] |

*Notation: () represents rank and \* represents proposed method*

To elaborate further with results, the histograms in fig. 4.23 were used to showcase and present the results virtually. The diagram revealed that the bars under the Neyman smooth test is higher than any other methods, and this suggest the good performance of the methods over any other considered methods including the existing ones.

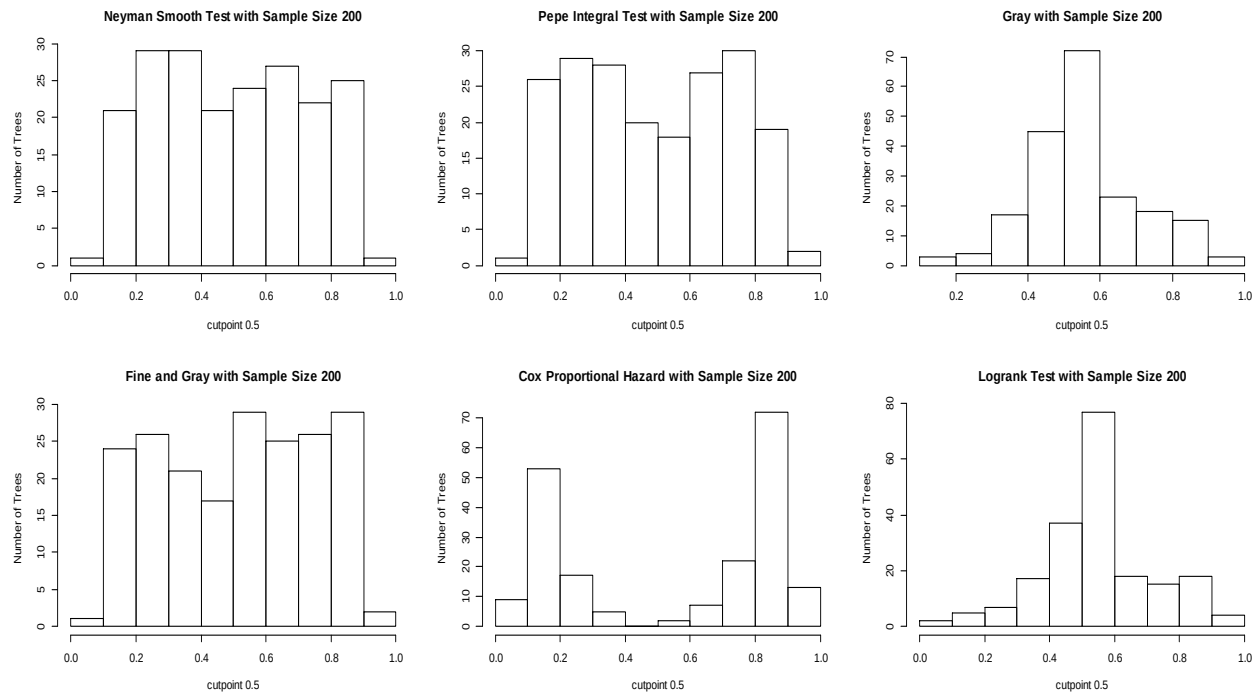


Fig. 4.23: Histogram of estimated cutpoint in exponential distribution

The results in Table 4.13 were further summarized using the relative performance indices. This was achieved by counting the number of times a particular method was ranked first and then

dividing it by the total number of sample sizes considered. This was done repeatedly until all the proposed and the existing methods were exhausted. The results were given in percentages for quick and easy comparison; all the corresponding results were presented in Table 4.14. Consequently, the results in table 4.14 revealed that: the proposed diagnostic classification tree using Neyman smooth as test statistic was found to yield a result that is better than any other method, across the eight selected sample sizes. In this case, Neyman smooth detect cutpoint better than any other existing methods since it has the highest relative performance of (50% and 75% for mean and median) than any other methods under exponential distribution. Considerably, under the mean the next best method is Pepe integral, but when we consider the median the Gray test was found to be the next best method to Neyman smooth test.

Table 4.14: Relative performance of the proposed and the existing methods across the sample sizes

| Relative performance (in %) |                 |            |            |                   |            |            |            |
|-----------------------------|-----------------|------------|------------|-------------------|------------|------------|------------|
| Measures                    | $T^{NS*}$       | $T^{KS*}$  | $T^{Ppe*}$ | $T^{G(existing)}$ | $T^{FG*}$  | $T^{CPH*}$ | $T^{LR}$   |
| Mean                        | <b>4/8 (50)</b> | 1/8 (12.5) | 2/8 (25)   | 0 (0)             | 1/8 (12.5) | 0 (0)      | 0 (0)      |
| Median                      | <b>6/8 (75)</b> | 0 (0)      | 0 (0)      | 1/8 (12.5)        | 0 (0)      | 0 (0)      | 1/8 (12.5) |

**Notation: () represents rank and \* represents proposed method**

The results in Table 4.15 revealed the amount of time (seconds) spent by the selected methods before the results were generated. It measures the efficiency of the test and their relative performance in term of the system time of any test statistic. It was observed from the results that as the sample size increases the amount of time needed to run any of the selected tests increases. Interestingly, Pepe was found to be highly effective in terms of system time and Neyman Smooth which was the best in terms of the median as was shown earlier, was found to be the second best in terms system time. However, the existing method was found to be in the fourth position, whereas,

it was in the low rank when the sample size is small and later improved as the sample sizes increases.

Table 4.15: Performance of methods based on system time (Exponential Distribution)

| Sample size | System time in Seconds |                       |                      |                      |                       |                      |                      |
|-------------|------------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|
|             | $T^{NS*}$              | $T^{KS*}$             | $T^{Pepe*}$          | $T^{G(existing)}$    | $T^{FG*}$             | $T^{CPH*}$           | $T^{LR}$             |
| 20          | 26.90 <sup>(2)</sup>   | 316.91 <sup>(6)</sup> | 10.05 <sup>(1)</sup> | 51.00 <sup>(4)</sup> | 539.97 <sup>(7)</sup> | 60.85 <sup>(5)</sup> | 47.68 <sup>(3)</sup> |
| 50          | 47.09                  | 822.67                | 18.83                | 84.48                | 791.88                | 175.20               | 121.48               |
| 80          | 83.69                  | 2014.86               | 25.79                | 127.08               | 1204.59               | 243.21               | 183.85               |
| 100         | 172.66                 | 2772.20               | 36.65                | 207.97               | 2464.57               | 293.68               | 219.26               |
| 150         | 194.28                 | 3232.75               | 51.49                | 287.63               | 3273.73               | 493.39               | 413.16               |
| 200         | 307.50                 | 5110.50               | 79.43                | 499.71               | 4141.25               | 620.95               | 455.78               |
| 500         | 763.60                 | 15378.17              | 221.11               | 1250.94              | 12938.86              | 1524.17              | 1074.31              |
| 1000        | 1100.67                | 19022.17              | 413.19               | 1643.97              | 18593.90              | 2905.64              | 2246.33              |

Notation: () represents rank and \* represents proposed method

## ii. Results of Log-normal Distribution on Between-Node Heterogeneity (Task 2)

The results of log-normal distribution under various sample sizes with 50% censoring rate were provided in detail in Table 4.16. The results show the various performances of the proposed diagnostic classification tree and the existing methods. This result is similar to the results in Table 4.16, and we observed that the proposed methods with Neyman Smooth test found to be the best among other methods. Moreover, the results also revealed the outstanding performance of the technique (Neyman Smooth test) in respect to the cutpoint detection due to the closeness of its estimate to the chosen cutpoint (0.5). The existing method was also found to be in the lowest rank and behaved equally as it was in the earlier results. Conclusively, the Neyman Smooth test adopted to account for the between node heterogeneity under log-normal and

exponential distributions was found to perform better than other proposed methods including the existing ones.

Table 4.16: Simulation with Log-normal Distribution over 50% censoring rate and 0.5 cut-points

| Sample size | Performance Measures | Two-sample Tests of the Equality of Two Cumulative Incidence Functions |                      |                      |                      |                      |                      |
|-------------|----------------------|------------------------------------------------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|             |                      | $T^{NS*}$                                                              | $T^{KS*}$            | $T^{Pepe*}$          | $T^{G(existing)}$    | $T^{FG*}$            | $T^{CPH*}$           |
| 20          | Mean                 | 0.4294(5)<br>[0.071]                                                   | 0.5485(3)<br>[0.049] | 0.5000(1)<br>[0.000] | 0.5682(4)<br>[0.068] | 0.5955(6)<br>[0.096] | 0.5009(2)<br>[0.001] |
|             | Median               | 0.4619(4)<br>[0.038]                                                   | 0.5201(3)<br>[0.020] | 0.4999(1)<br>[0.000] | 0.6231(5)<br>[0.123] | 0.7197(6)<br>[0.220] | 0.5009(2)<br>[0.001] |
| 50          | Mean                 | 0.4752(3)<br>[0.025]                                                   | 0.5289(4)<br>[0.029] | 0.4998(1)<br>[0.000] | 0.5433(5)<br>[0.043] | 0.5914(6)<br>[0.091] | 0.5007(2)<br>[0.001] |
|             | Median               | 0.4869(4)<br>[0.013]                                                   | 0.4946(3)<br>[0.005] | 0.5001(1)<br>[0.000] | 0.6018(5)<br>[0.102] | 0.6396(6)<br>[0.140] | 0.5005(2)<br>[0.001] |
| 80          | Mean                 | 0.5004(1)<br>[0.000]                                                   | 0.5280(4)<br>[0.028] | 0.4998(1)<br>[0.000] | 0.5261(3)<br>[0.026] | 0.5590(5)<br>[0.059] | 0.5005(2)<br>0.001   |
|             | Median               | 0.4998(1)<br>[0.000]                                                   | 0.5017(2)<br>[0.002] | 0.4999(1)<br>[0.000] | 0.5173(3)<br>[0.017] | 0.6204(4)<br>[0.120] | 0.5002(1)<br>[0.000] |
| 100         | Mean                 | 0.4875(3)<br>[0.013]                                                   | 0.5241(4)<br>[0.024] | 0.5003(1)<br>[0.000] | 0.5292(5)<br>[0.029] | 0.5723(6)<br>[0.072] | 0.5006(2)<br>[0.001] |
|             | Median               | 0.4869(3)<br>[0.013]                                                   | 0.5017(2)<br>[0.002] | 0.5002(1)<br>[0.000] | 0.5340(4)<br>[0.034] | 0.6207(5)<br>[0.121] | 0.5004(1)<br>[0.000] |
| 150         | Mean                 | 0.4908(3)                                                              | 0.5247(5)            | 0.5003(1)            | 0.5165(4)            | 0.5836(6)            | 0.5009(2)            |

|      |        |                              |                              |                              |                              |                              |                              |
|------|--------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
|      |        | [0.009]                      | [0.025]                      | [0.000]                      | [0.017]                      | [0.084]                      | [0.001]                      |
|      | Median | 0.5025 <b>(4)</b><br>[0.003] | 0.5017 <b>(3)</b><br>[0.002] | 0.5003 <b>(1)</b><br>[0.000] | 0.5340 <b>(5)</b><br>[0.034] | 0.6174 <b>(6)</b><br>[0.117] | 0.5005 <b>(2)</b><br>[0.001] |
| 200  | Mean   | 0.4809 <b>(4)</b><br>[0.019] | 0.5247 <b>(5)</b><br>[0.025] | 0.5003 <b>(1)</b><br>[0.000] | 0.5152 <b>(3)</b><br>[0.015] | 0.5583 <b>(6)</b><br>[0.058] | 0.5008 <b>(2)</b><br>[0.001] |
|      | Median | 0.4876 <b>(4)</b><br>[0.012] | 0.5017 <b>(3)</b><br>[0.002] | 0.5004 <b>(1)</b><br>[0.000] | 0.5427 <b>(5)</b><br>[0.043] | 0.6019 <b>(6)</b><br>[0.102] | 0.5006 <b>(2)</b><br>[0.001] |
| 500  | Mean   | 0.4906 <b>(5)</b><br>[0.009] | 0.5041 <b>(3)</b><br>[0.004] | 0.5002 <b>(1)</b><br>[0.000] | 0.4945 <b>(4)</b><br>[0.006] | 0.5548 <b>(6)</b><br>[0.055] | 0.5008 <b>(2)</b><br>[0.001] |
|      | Median | 0.5048 <b>(4)</b><br>[0.005] | 0.4980 <b>(2)</b><br>[0.002] | 0.5000 <b>(1)</b><br>[0.000] | 0.4962 <b>(3)</b><br>[0.004] | 0.6165 <b>(5)</b><br>[0.117] | 0.5002 <b>(1)</b><br>[0.000] |
| 1000 | Mean   | 0.4953 <b>(2)</b><br>[0.005] | 0.5094 <b>(3)</b><br>[0.009] | 0.5005 <b>(1)</b><br>[0.001] | 0.4972 <b>(4)</b><br>[0.003] | 0.5720 <b>(5)</b><br>[0.072] | 0.5011 <b>(1)</b><br>[0.001] |
|      | Median | 0.5030 <b>(3)</b><br>[0.003] | 0.5090 <b>(5)</b><br>[0.009] | 0.5002 <b>(1)</b><br>[0.000] | 0.5070 <b>(4)</b><br>[0.007] | 0.6284 <b>(6)</b><br>[0.128] | 0.5005 <b>(2)</b><br>[0.001] |

*Notation: () represents rank and \* represents proposed method*

The study further displays the results in table 4.16 using the histograms in fig. 4.24. The diagram revealed the performance of the proposed model over the existing techniques since the bars under the Neyman smooth test were almost on the same height than any other methods, and this suggest the good performance of the methods over any other considered methods including the existing ones. This diagram is similar to the one presented in fig. 4.24, thus can be interpreted the same way.

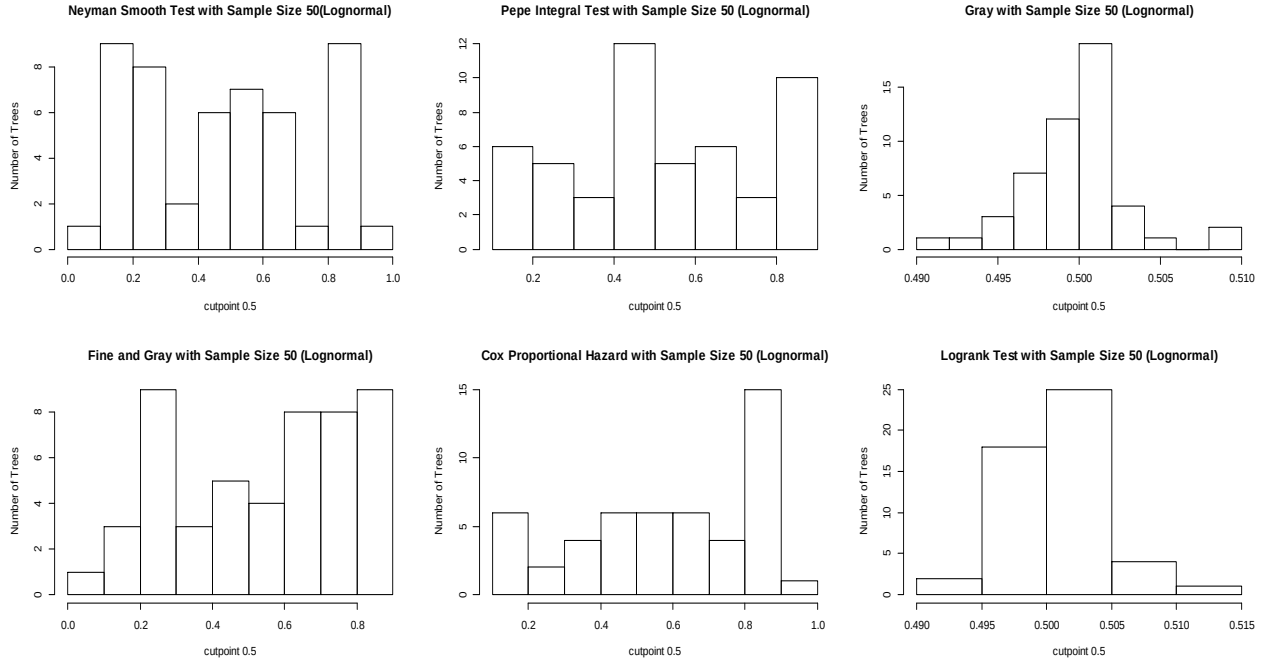


Fig. 4. 24: Histogram of estimated cutpoint in Log-normal distribution

The results in Table 4.16 were further summarized in Table 4.17 in order to see the behavior of the proposed model under lognormal. Surprisingly, it was observed that the existing method Gray perform better than the proposed methods with mean and median of 50% and 50% against 37.5% and 12.5% for Neyman Test. However, we found that the proposed models outperform the existing model at varying sample sizes as shown in Table 4.17.

Table 4.17: Relative performance of the proposed and the existing methods across the sample sizes

| Relative performance (in %) (Lognormal Distribution) |            |            |                   |           |            |            |
|------------------------------------------------------|------------|------------|-------------------|-----------|------------|------------|
| Measures                                             | $T^{NS*}$  | $T^{Ppe*}$ | $T^{G(existing)}$ | $T^{FG*}$ | $T^{CPH*}$ | $T^{LR}$   |
| Mean                                                 | 3/8 (37.5) | 0(0)       | 04/8(50)          | 0(0)      | 0 (0)      | 1/8 (12.5) |
| Median                                               | 1/8 (12.5) | 1/8(12.5)  | 4/8 (50)          | 0 (0)     | 0 (0)      | 2/8 (25)   |

For clarity purpose, the system time for the log-normal was also considered, and it was observed from Table 4.18 that the proposed methods Pepe and Neyman Smooth were found to be the most

efficient techniques in term of system time. Gray technique was found to be the third best and cox proportional hazard relatively behaves better.

In term of categorizing a continuous covariate, there is very little attention given to this aspect in biostatistics, epidemiology, engineering and other fields. Our research work currently provides an insight into cutpoint determination methods in classification and regression tree, and to be specific in competing risk model. This gives the wide range applicability of our proposed diagnostic classification tree.

Table 4.18: Performance of methods based on system time (Lognormal Distribution)

| Sample size | System time in Seconds |                       |                     |                      |                       |                      |                      |
|-------------|------------------------|-----------------------|---------------------|----------------------|-----------------------|----------------------|----------------------|
|             | $T^{NS*}$              | $T^{KS*}$             | $T^{Pepe*}$         | $T^{G(existing)}$    | $T^{FG*}$             | $T^{CPH*}$           | $T^{LR}$             |
| 20          | 16.36 <sup>(2)</sup>   | 238.35 <sup>(6)</sup> | 9.62 <sup>(1)</sup> | 23.87 <sup>(3)</sup> | 354.73 <sup>(7)</sup> | 24.26 <sup>(4)</sup> | 32.86 <sup>(5)</sup> |
| 50          | 36.16                  | 838.06                | 18.45               | 57.96                | 529.40                | 66.44                | 77.14                |
| 80          | 54.93                  | 1510.71               | 32.74               | 82.03                | 707.20                | 91.19                | 123.96               |
| 100         | 70.16                  | 1711.24               | 44.50               | 117.36               | 1561.22               | 114.97               | 150.44               |
| 150         | 103.50                 | 3778.89               | 71.67               | 313.75               | 1955.79               | 219.50               | 227.28               |
| 200         | 142.27                 | 3384.07               | 79.87               | 419.90               | 2058.34               | 260.31               | 244.82               |
| 500         | 340.48                 | 6875.66               | 142.75              | 996.54               | 4879.08               | 822.43               | 592.72               |
| 1000        | 1027.34                | 10939.13              | 413.36              | 2052.64              | 9491.19               | 1470.34              | 1249.57              |

Notation: () represents rank and \* represents proposed method

#### 4.2.1 Application of survival based-tree via between-node heterogeneity on Real-life Dataset

The proposed between-node heterogeneity methods were used on follicular cell lymphoma data recorded on the number of ovarian follicles, on a log scale, counted from sectioned ovaries of women of various ages.

The follicular cell lymphoma study was introduced in the package “randomSurvivalForest” through the language R, and it contains two competing risks (relapse and death) with two covariates: age and Hemoglobin (g/l).

Results in Table 4.19 provide the summary statistics of age and hemoglobin (g/l) for all the Clinical stages in follicular cell lymphoma. The median or the mean of age and hemoglobin can be used as cutpoint detection in a simple data-oriented methodology. Thus, the aim of this research is to apply Classification Tree (CART) to survival analysis with two risk factors and the results of cutpoint generated for various statistic adopted were presented.

Table 4.19: Summary Statistics of covariates Age and hemoglobin (g/l)

| <b>Age</b>              |          |        |       |           |       |       |
|-------------------------|----------|--------|-------|-----------|-------|-------|
| Group                   | <i>n</i> | Median | Mean  | Std. Dev. | Min   | Max   |
| Clinical stage I        | 362      | 60.00  | 58.32 | 14.06372  | 17.00 | 86.00 |
| Clinical stage II       | 179      | 56.00  | 54.21 | 13.89789  | 22.00 | 79.70 |
| <b>Hemoglobin (g/l)</b> |          |        |       |           |       |       |
| Clinical stage I        | 362      | 140.0  | 139.3 | 14.97512  | 40.0  | 189.0 |
| Clinical stage II       | 179      | 140.0  | 137.3 | 16.05435  | 80.0  | 172.0 |

The various results of the proposed and existing models were presented in table 4.20. The table was formulated by taking the absolute differences between the estimated cutpoint and the true median. The results of the absolute differences were later ranked from the lowest to the highest. We observed that the proposed adopted methods Pepe and Fine and Gray were found to be the best compared to any other adopted methods including existing methods based on the covariate age. Interestingly, under the covariate hemoglobin level, the Cox-proportional hazard, Fine and Gray outperform any other methods used in this research.

Table 4.20: Maximum cutpoint generated across the selected test

|                  | <b>Age</b>              |          |            |                     |                 |                 |          |
|------------------|-------------------------|----------|------------|---------------------|-----------------|-----------------|----------|
| Group            | $T^{NS}$                | $T^{KS}$ | $T^{Pepe}$ | $T^{G(fiona,2008)}$ | $T^{FG}$        | $T^{CPH}$       | $T^{LR}$ |
| Cutpoint         | 32.1000                 | 32.1000  | <b>52</b>  | 35.5500             | <b>73.4000</b>  | 32.1000         | 37.9000  |
| $ Difference $ * | 27.9[5]                 | 27.9[5]  | 8 [1]      | 24.45[4]            | 13[2]           | 27.9[5]         | 22.1[3]  |
|                  | <b>Hemoglobin (g/l)</b> |          |            |                     |                 |                 |          |
| Cutpoint         | 111.5000                | 111.5000 | 145.5000   | 146.5000            | <b>143.0000</b> | <b>139.5000</b> | 145.5000 |
| $ Difference $ * | 28.5[5]                 | 28.5[5]  | 5.5[3]     | 6.5[4]              | 3[2]            | 0.5[1]          | 5.5[3]   |

\*Is the absolute difference between the true median and the cutpoint and [] is the rank value

#### 4.2.2 Estimation of Optimal cutpoint on between-node heterogeneity

In this section, we divert our attention to a method proposed by Contal and Quigley (1999), which is used to test the significance of the chosen cutpoint based on the continuous covariates only.

We apply this approach to a real-life data on follicular cell lymphoma mentioned in section 3.2.5, the full results of the method were presented in appendix A based on age and Hemoglobin (g/l). The potential cutpoint of age selected by this method is 57.4years (Appendix A1) with  $p$ -value( $<0.05$ ) and this suggests that the cutpoint obtained is significant and thus age is related to the survival of the patient from the disease. The second covariate Hemoglobin has a cutpoint of 140g/l (Appendix A2) with  $p$ -value ( $>0.05$ ) that suggests non-significant of the Hemoglobin and fails to relate to the survival of patient from the disease.

Comparably, the cutpoint of age chosen by the proposed diagnostics classification tree by pepe is 52 years (Table 4.20), and it is the closest cutpoint to the 57.4 years than any other proposed and existing methods. Additionally, the Cox-PH was found to have a closer cutpoint 139g/l to 140g/l than any other considered techniques. These support the performances of the proposed methods over the existing methods in terms of cutpoint detection.

### **4.3 Results of Artificial Neural Network in Survival Analysis with Competing Risks (Task 3)**

In this research work, three different real-life datasets were considered and used to validate the performance of the proposed model against the classical model. The first dataset was on double-blinded randomized clinical trial, conducted in order to compare two treatments for patients with prostate cancer. This data set was collected from Royal Berkshire Hospitals, Reading, United Kingdom, and is freely available online, please see additional material for full description. (Please kindly see the attached CD for full description of the data).

The main aims of this clinical trial experiment were to determine how survival time of patients was affected by the treatment and any other prognostic factor(s). Also, possible toxic effect of DES particularly in terms of mortality from heart disease was to be examined. However, endpoint considered was death from prostatic cancer and death from any other causes. The following covariates were used for this research, others were discarded due to missing observations: Age in years, Weight index (WTINDEX), Performance rating (PERF), Cardiovascular Disease History (HISTCD), Electrocardiogram Code (ECGCODE), Tumour Size (SIZEPT), Tumour Stage&Grade Index (SHINDEX) and Serum Haemoglobin gm/100ml (SERUMHG). The Follow-up Time is in Months and the Survival Status was recorded by treating death from prostatic cancer as the event of interest (cause 1), death from Heart or Vascular Disease and other causes as competing risks (cause 2) and discharge alive as true censoring (event 0).

The second dataset was on melanoma data frame with 205 rows and 7 columns. It contains data relating to survival of patients after operation for malignant melanoma collected at Odense University Hospital by K.T. Drzewiecki (Andersen et al., 1993).

The third dataset was on Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. There is a lot of missing observation in the dataset and this was taken care off by EM-algorithm.

#### 4.3.1 Comparing the Proposed Model with Classical Competing Risk Model

The comparison between the proposed and classical competing risk model was assessed using the Root Mean Square Error (RMSE). The RMSE presents information on the short-term efficiency which is the difference between the predicated values and the observed values. The lower the RMSE, the more accurate the evaluation. The second criterion is Mean Absolute Error (MAE), which is the average deviation of the predicted values from the corresponding observed values. The lower the MAE, the better the long-term model prediction. Finally, the Relative Squared Error (RSE) criterion was also adopted and is the aggregate squared error produce relative to what the error would have been if the prediction had been the average of the absolute value. The lower the RSE, the better model prediction.

Table 4.21: Double-blinded randomized clinical trial with first and second events (death due to Prostatic Cancer and Other diseases)

| <b>Event 1: Death due to Prostatic Cancer</b> |               |               |               |
|-----------------------------------------------|---------------|---------------|---------------|
| Model                                         | RMSE          | MAE           | RSE           |
| <b>Proposed model (CMLP)</b>                  | <b>0.1237</b> | <b>0.1397</b> | <b>0.9933</b> |
| <b>Classical model</b>                        | 0.3020        | 0.1709        | 4.6391        |
| <b>Event 2: Death due to Other Diseases</b>   |               |               |               |
| <b>Proposed model (CMLP)</b>                  | <b>0.2021</b> | <b>0.2519</b> | <b>1.0000</b> |
| <b>Classical model</b>                        | 0.2261        | 0.2763        | 1.2036        |

The results of death due to prostate cancer for both proposed and existing models were presented in Table 4.21 and it was observed that irrespective of the criteria, the proposed model (CMPL) performed better than the existing technique. Additionally, the result of deaths due to other diseases also indicates good performance of proposed model over the classical methods.

Table 4.22: Melanoma skin cancer incidence with first and second events (death from melanoma and other causes)

| <b>Event 1: Death from melanoma</b>       |               |               |               |
|-------------------------------------------|---------------|---------------|---------------|
| Model                                     | RMSE          | MAE           | RSE           |
| <b>Proposed model (CMLP)</b>              | <b>0.0921</b> | <b>0.1132</b> | <b>0.8092</b> |
| <b>Classical model</b>                    | 0.1961        | 0.2766        | 4.8318        |
| <b>Event 2: Death due to other causes</b> |               |               |               |
| <b>Proposed model (CMLP)</b>              | 0.0288        | <b>0.0375</b> | <b>0.9404</b> |
| <b>Classical model</b>                    | <b>0.0248</b> | 0.0387        | 1.0001        |

Death due to melanoma results were presented in table 4.22 and we observed that the proposed model (CMLP) fit the dataset better than the existing model irrespective of the criteria. Moreover, the results under death due to other causes revealed that the proposed model performed better than the existing model in MAE and RSE but failed when RMSE was considered.

Table 4.23: Mayo Clinic Primary Biliary Cirrhosis Data with first and second events (death from Transplant and Remission)

| <b>Event 1: Death due to Transplant</b> |               |               |               |
|-----------------------------------------|---------------|---------------|---------------|
| Model                                   | RMSE          | MAE           | RSE           |
| <b>Proposed model (CMLP)</b>            | <b>0.0335</b> | <b>0.0391</b> | <b>0.9992</b> |
| <b>Classical model</b>                  | 0.0687        | 0.1276        | 10.6277       |
| <b>Event 2: Death due to Remission</b>  |               |               |               |
| <b>Proposed model (CMLP)</b>            | <b>0.2019</b> | <b>0.2614</b> | <b>0.9997</b> |
| <b>Classical model</b>                  | 0.3408        | 0.5979        | 5.2290        |

Finally, the third dataset considered is on primary biliary cirrhosis (PBC) of the liver. It also consists two competing risks and seventeen (17) covariates, the risks are; death due to transplant and death in remission. Table 4.23 presents the results of death due to transplant risk, this result is similar to the previous results, by showing the efficiency of the proposed model over the existing model irrespective of the criteria. The results of death in remission in table 6 shows different results when compared with the previous dataset, it revealed that the proposed model in the second competing risks performed better than the existing model irrespective of any criteria.

#### 4.2.3 Implementation CMLPANN on Double-Blinded Randomized Clinical Data

Further analysis was carried out on Double-blinded randomized clinical trial data and this analysis includes; the effect of hidden layers and the interpretation of results.

The results of the proposed model (CMLPANN) were presented based on the various levels of the competing events in the dataset. The first risk event is death due to Prostatic Cancer (*i.e.*  $m = 1$ ) and the second risk is death due to other diseases (*i.e.*  $m = 2$ ). The various hidden layer results

were presented in Table 4.24 and the sum of squared error indices was used to check the performance of the best hidden layer using resilient back propagation.

Table 4.24: Resilient back propagation with weight for event1 (death due to Prostatic Cancer)

| Hidden layers | Sum of Squared Error | Reached Threshold | Steps |
|---------------|----------------------|-------------------|-------|
| 2             | 2.3196               | 0.00001           | 22    |
| 5             | 2.3128               | 0.007             | 24    |
| 10            | 1.8250               | 0.0093            | 4277  |
| 15            | 1.2254               | 0.0097            | 40391 |

Table 4.24 shows the results of CMLP model with four different hidden layers, the hidden layers were 2, 5, 10 and 15. It can be deduced from the results that as the hidden layers increase the sum of square error reduces, threshold increases and steps increase respectively. Meanwhile, the threshold is an integer for specifying the partial derivatives of the error function for stopping criteria, as discussed in the proposed model methodology. Likewise, the step is the number of repetitions for the training process (Frauk et al., 2010). The only effect of the hidden layers is on the number steps before the result converge, the result revealed the high number of steps before conversion due to high use of hidden layers.

The virtual presentation of the complete proposed CMLP model that include the input and output variables were given in fig. 4.25 and the results of training process with 2 hidden layers were showcased. The fig. 4.25 at the same time displays the trained synaptic weights, different intercept and basic information about the training process like the overall error to mimic human brain as explained earlier in chapter one.

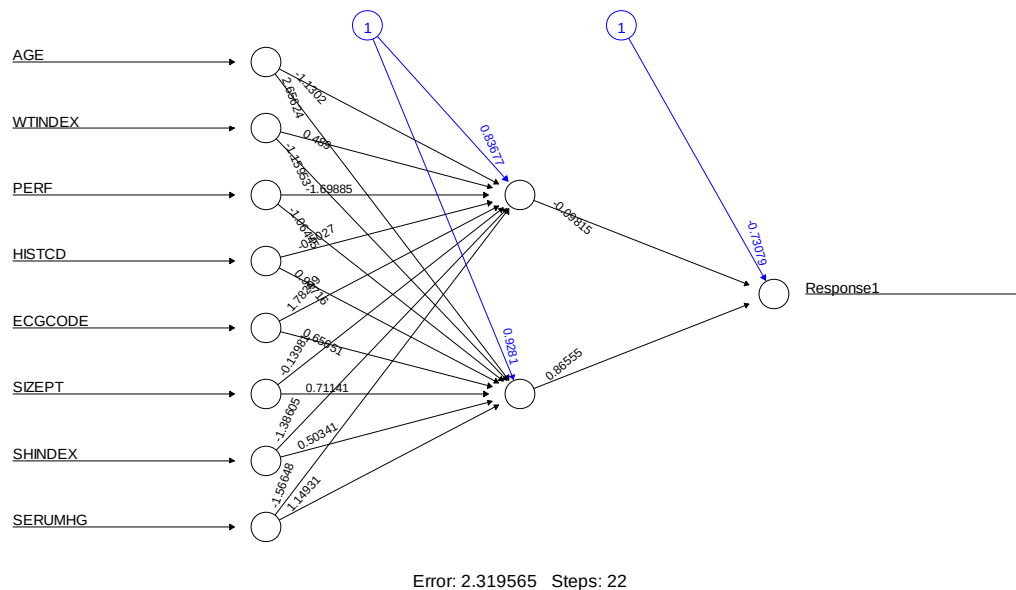


Fig. 4.25: Plot of a trained neural network including trained synaptic weights and basic information about the training process.

The relative important values in Table 4.25 were estimated using the methods of Garson (1991) as explained above and its results was presented graphically in Fig. 4.2. The results tell us that the covariates Age, Tumour Stage, Grade Index (SHINDEX) and Electrocardiogram Code (ECGCODE) have the strongest positive and negative relationship, individually, on the response (probability of survival) variable. Correspondingly, covariates that have relative importance close to zero, such as Weight Index (WTINDEX), do not have any substantial importance on response variable. Nevertheless, this covariate with zero relative importance most likely have some marginal effect on the response variable, but its effect is irrelevant in the context of the other covariates. Other covariates were found to be either very weak negative and positive relationship with the response variable as shown in table 4.25 and fig. 4.26.

Table 4.25: Variable Important by Garson Algorithm (1991)

| Covariate                               | Relative Important | Remarks on the survival of the patient |
|-----------------------------------------|--------------------|----------------------------------------|
| Age in years (AGE)                      | 0.2268             | Evidence of positive relationship      |
| Weight Index (WTINDEX)                  | 0.0001             | No importance                          |
| Performance Rating (PERF)               | -0.4577            | Fair negative relationship             |
| Cardiovascular Disease History (HISTCD) | -0.4635            | Fair negative relationship             |
| Electrocardiogram Code (ECGCODE)        | -1.0000            | Perfect negative relationship          |
| Tumour Size(SIZEPT)                     | -0.0397            | Weak negative relationship             |
| Tumour Stage, Grade Index (SHINDEX)     | -0.7489            | Strong negative relationship           |
| Serum Haemoglobin gm/100ml (SERUMHG)    | 0.13292            | Weak positive relationship             |

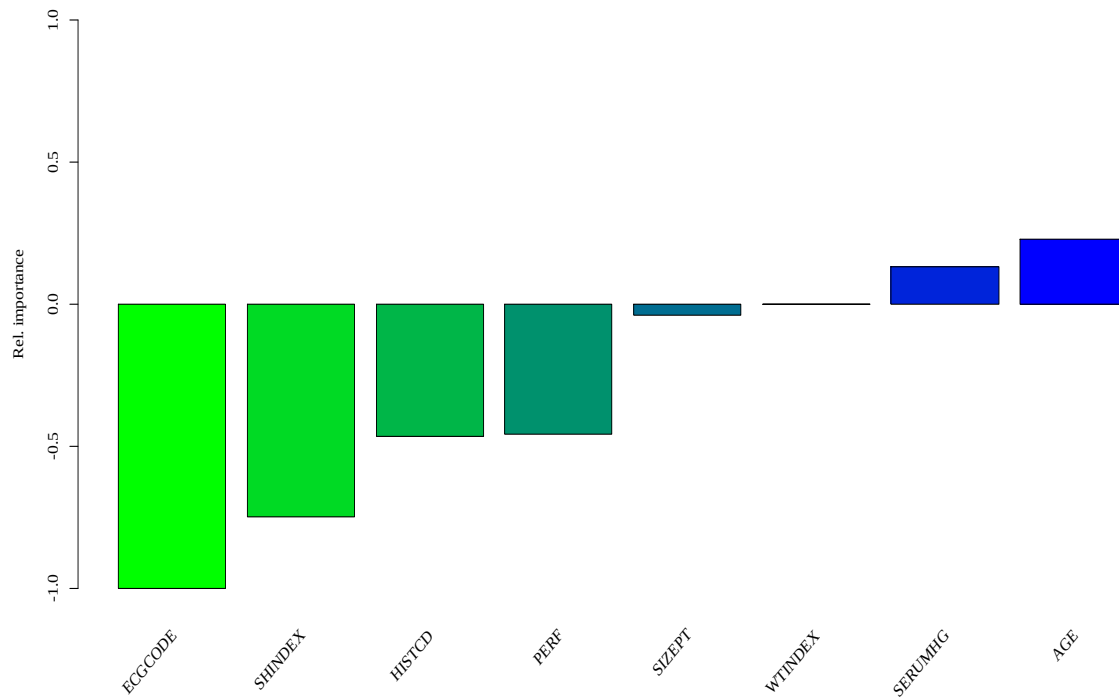


Fig. 4.26: Relative importance of the eight covariates for response (probability of survival) variable using CMLP Model

#### 4.4 Results of Multivariate Adaptive Regression Splines in Survival Analysis (Task 4)

In this section, the proposed methodology (MARS) was applied to both simulation study explained in chapter four and the real-life data. The various generated results on simulation and real-life data were presented in the following subsections.

##### 4.4.1 Results of the Simulation on Survival MARS

In the Monte-Carlo experiment, 1000 simulation cycles were computed for four different values of the penalizing parameter  $d$ ,  $d: \in \{2, 3, 4, 5\}$ , with martingale and Cox-Snell residuals as responses. Also, three different censoring rates of low (25%), moderate (50%) and high (85%) were considered in the analysis.

At each fit of survival MARS model on the simulated data, the amount of non-linear influence of the covariate  $x_1$  detected by the models (expressed in %) as well as that of its interaction with  $x_2$  at different values of the penalizing parameters were determined under the Cox-Snell and Martingale residual responses.

Also, the amount of false basis function (in %) chosen by each model was also determined under the two residual types.

Tables 4.26, 4.27 and 4.28 presented the results of the simulation studies under the three selected distributions employed for simulating the survival times.

Table 4.26: Results of Survival MARS with Exponential distributed survival time. The amount of non-linear effects of covariates  $x_1$  (in %) and its interactions with  $x_2$  with Martingale and Cox-Snell residuals as responses were reported.

| d | Martingale Residuals |                 |            |              | Cox-Snell Residuals |                 |            |              |
|---|----------------------|-----------------|------------|--------------|---------------------|-----------------|------------|--------------|
|   | Censoring rate       | %knots( $x_1$ ) | % $x_1x_2$ | %false basis | Censoring rate      | %knots( $x_1$ ) | % $x_1x_2$ | %false basis |
| 2 | 25%                  | 70%             | 100%       | 50%          | 25%                 | 80%             | 100%       | 50%          |
|   | 50%                  | 100%            | 100%       | 0%           | 50%                 | 100%            | 100%       | 30%          |
|   | 85%                  | 100%            | 80%        | 70%          | 85%                 | 90%             | 30%        | 80%          |
| 3 | 25%                  | 70%             | 100%       | 40%          | 25%                 | 80%             | 100%       | 40%          |
|   | 50%                  | 100%            | 100%       | 0.0%         | 50%                 | 100%            | 100%       | 0.0%         |
|   | 85%                  | 100%            | 50%        | 60%          | 85%                 | 100%            | 100%       | 20%          |
| 4 | 25%                  | 40%             | 100%       | 60%          | 25%                 | 60%             | 100%       | 50%          |
|   | 50%                  | 100%            | 100%       | 0.0%         | 50%                 | 100%            | 90%        | 30%          |
|   | 85%                  | 100%            | 80%        | 30%          | 85%                 | 50%             | 0.00%      | 100%         |
| 5 | 25%                  | 20%             | 100%       | 80%          | 25%                 | 50%             | 100%       | 50%          |
|   | 50%                  | 90%             | 100%       | 10%          | 50%                 | 100%            | 100%       | 0.0%         |
|   | 85%                  | 100%            | 30%        | 70%          | 85%                 | 50%             | 0.0%       | 100%         |

Table 4.27: Results of the simulation studies on Weibull distributed ( $\nu = 2, \lambda = 5$ ) survival time with Martingale and Cox-Snell residuals.

| d | Martingale Residuals |                 |            |              | Cox-Snell Residuals |                 |            |              |
|---|----------------------|-----------------|------------|--------------|---------------------|-----------------|------------|--------------|
|   | Censoring rate       | %knots( $x_1$ ) | % $x_1x_2$ | %false basis | Censoring rate      | %knots( $x_1$ ) | % $x_1x_2$ | %false basis |
| 2 | 25%                  | 80%             | 100%       | 50%          | 25%                 | 100%            | 100%       | 40%          |
|   | 50%                  | 100%            | 100%       | 0%           | 50%                 | 100%            | 100%       | 50%          |
|   | 85%                  | 100%            | 50%        | 70%          | 85%                 | 100%            | 50%        | 60%          |
| 3 | 25%                  | 40%             | 100%       | 60%          | 25%                 | 100%            | 100%       | 0%           |
|   | 50%                  | 100%            | 100%       | 10%          | 50%                 | 100%            | 90%        | 30%          |
|   | 85%                  | 100%            | 50%        | 60%          | 85%                 | 100%            | 60%        | 60%          |
| 4 | 25%                  | 10%             | 100%       | 90%          | 25%                 | 100%            | 100%       | 50%          |
|   | 50%                  | 100%            | 90%        | 20%          | 50%                 | 100%            | 80%        | 50%          |
|   | 85%                  | 90%             | 30%        | 80%          | 85%                 | 90%             | 40%        | 50%          |
| 5 | 25%                  | 0%              | 100%       | 100%         | 25%                 | 100%            | 100%       | 0%           |
|   | 50%                  | 100%            | 100%       | 30%          | 50%                 | 100%            | 60%        | 40%          |
|   | 85%                  | 100%            | 40%        | 60%          | 85%                 | 100%            | 50%        | 50%          |

Table 4.28: Results of the simulation studies on Gompertz distributed survival time with Martingale and Cox-Snell residuals.

| d | Martingale Residuals |                 |            |              | Cox-Snell Residuals |                 |            |              |
|---|----------------------|-----------------|------------|--------------|---------------------|-----------------|------------|--------------|
|   | Censoring rate       | %knots( $x_1$ ) | % $x_1x_2$ | %false basis | Censoring rate      | %knots( $x_1$ ) | % $x_1x_2$ | %false basis |
| 2 | 25%                  | 60%             | 80%        | 50%          | 25%                 | 80%             | 60%        | 70%          |
|   | 50%                  | 90%             | 30%        | 80%          | 50%                 | 90%             | 50%        | 70%          |
|   | 85%                  | 70%             | 10%        | 100%         | 85%                 | 90%             | 30%        | 80%          |
| 3 | 25%                  | 60%             | 100%       | 70%          | 25%                 | 90%             | 40%        | 80%          |
|   | 50%                  | 60%             | 60%        | 70%          | 50%                 | 80%             | 10%        | 90%          |
|   | 85%                  | 80%             | 40%        | 60%          | 85%                 | 80%             | 0%         | 100%         |
| 4 | 25%                  | 40%             | 80%        | 80%          | 25%                 | 70%             | 30%        | 100%         |
|   | 50%                  | 30%             | 10%        | 90%          | 50%                 | 80%             | 0%         | 100%         |
|   | 85%                  | 80%             | 0%         | 100%         | 85%                 | 60%             | 20%        | 90%          |
| 5 | 25%                  | 30%             | 80%        | 70%          | 25%                 | 60%             | 20%        | 80%          |
|   | 50%                  | 50%             | 10%        | 90%          | 50%                 | 50%             | 0%         | 100%         |
|   | 85%                  | 50%             | 0%         | 10%          | 85%                 | 40%             | 0%         | 100%         |

The implications of the results in Table 4.26 regarding the behaviors of predictor variables  $x_1$  and  $x_2$  in the proposed (Cox-Snell) and the existing (Martingale) models under the Exponential survival times were clearly shown by the graphs (see Appendix B, Figs. 1 to 4) in Appendix B at the selected values of the penalty (tuning) parameter  $d = 2$  and  $5$  with low (25%) and high (85%) censoring rates.

It was found that both models identified the inherent linear relationship of the nuisance variables  $x_3, x_4$  and  $x_5$  with the response variable ( $t$ ). However, the structure displayed by the various graphs showed that the survival MARS model with Cox-Snell residual (proposed model) was able to capture the linear, nonlinear and interaction effects of the covariates  $x_1$  and  $x_2$  better than the survival MARS model with Martingale residuals (existing model) (see Appendix B, Figs. 1 to 4).

Furthermore, the graphical presentation of the results for Weibull distributed survival times was also presented in appendix C (Figs. 5 to 8). These graphical results equally provided information on the implication of the results in Table 3 at penalty parameter values  $d = 2$  and  $5$  with low (25%)

and high (85%) censoring rates. Results from the graphs (Figs 5 to 8) showed that both the Cox-Snell and Martingale models identified the nuisance variables  $x_3, x_4$  and  $x_5$  to be linearly concomitant to the response variable ( $t$ ), but the interaction effect of covariates  $x_1$  and  $x_2$  is tighter (under the Cox-Snell model) at large value of the penalty parameter ( $d = 5$ ) than at low values of  $d$  ( $d = 2$ ) especially at high censoring rate.

Finally, the results from the survival MARS under the two residual types with Gompertz distributed survival times followed similar patterns like those earlier presented. The graphical summary of the non-linear effects of the predictor variables  $x_1$  and  $x_2$  as well as their interactions were clearly shown by the graphs Appendix D (see Figs. 9 to 12). The linear association between other covariates  $x_3, x_4$  and  $x_5$  with survival time was also obvious from the various graphs.

#### **4.4.2 Application of Survival MARS to Pneumocystis Cariniip Neumonia Data**

To validate the results from the Monte-Carlo study presented in this section, the survival MARS models with Cox-Snell and Martingale residuals were both fitted to a published real-life dataset on Pneumocystis Cariniip Neumonia (PCP). These data had been previously analyzed by Chap (2003). The PCP, a life-threatening disease, has been reported to be the most common opportunistic infection in HIV-infected patients (Chap, 2003). Many North Americans with AIDS have had one or two episodes of PCP during the course of their HIV infection. The PCP has therefore become a factor to be considered in mortality and morbidity modeling while its recurrences are often common.

Table 4.29: Results of Survival MARS models using Cox-Snell and Martingale residuals on real life (PCP) data. Results of the classical Cox PH model fitted to these data were also reported in the table.

| Predictors                                        | Coefficient ( $\beta$ ) | Standard Error of $\beta$ |
|---------------------------------------------------|-------------------------|---------------------------|
| <b>Survival MARS with Martingale residual</b>     |                         |                           |
| CD4 count                                         | -0.00014                | 0.00025                   |
| Gender (Female = Ref.)                            | 0.15814                 | 0.10750                   |
| Race (white=1) (Other = Ref.)                     | 0.01463                 | 0.05310                   |
| Race (black=2) (Other = Ref.)                     | -0.01044                | 0.05880                   |
| Weight                                            | 0.00053                 | 0.00053                   |
| Homosexuality (No = Ref.)                         | -0.01291                | 0.04868                   |
| <b>Survival MARS with Cox-Snell residual</b>      |                         |                           |
| CD4 count                                         | 0.00005                 | 0.00005                   |
| Gender (Female = Ref.)                            | -0.01774                | 0.02511                   |
| Race (white=1) (Other = Ref.)                     | 0.02390                 | 0.01226                   |
| Race (black=2) (Other = Ref.)                     | -0.30309                | 0.01354                   |
| Weight                                            | -0.00006                | 0.00012                   |
| Homosexuality (No = Ref.)                         | 0.00422                 | 0.01133                   |
| <b>Classical Survival Analysis (Cox PH Model)</b> |                         |                           |
| CD4 count                                         | 0.99900                 | 0.00215                   |
| Gender (Female = Ref.)                            | $2.85 \times 10^{-7}$   | 3760                      |
| Race (white=1) (Other = Ref.)                     | 1.21000                 | 0.7480                    |
| Race (black=2) (Other = Ref.)                     | 0.98500                 | 0.84200                   |
| Weight                                            | 1.00000                 | 0.00434                   |
| Homosexuality (No = Ref.)                         | 0.73600                 | 0.37700                   |

The data consist of time to recurrence (in months) of PCP with four categorical predictors – gender (male, female), race (white, black, others), Homosexuality practice (yes, no) and two metrical (continuous) covariates - CD4 count and Weight. The outcome (status) variable is categorized as 1 if there was a relapse and 0 if no relapse occurred (censored). Results in Table 4.29 showed that the Survival MARS model with Cox-Snell residual performed better than the Martingale residual model using the standard errors of the estimated parameters as assessment criteria.

To allow broader comparison of the efficiency of the proposed survival MARS model using Cox-Snell residual with the classical Cox PH model, the Cox PH model was fitted to the data and the results obtained were equally reported in the Table 4.29. From the results in Table 4.29, it can be observed that the Cox-Snell residual model was more efficient than the Cox PH model on the real-life data as well. In the three cases, the standard errors of the estimated parameters by the Cox-Snell survival MARS model were relatively smaller than those provided by both the Cox PH and Martingale residual models.

## CHAPTER FIVE

### Summary, Conclusion and Recommendation

This chapter concludes this thesis. A summary of the research is presented, and findings of the study are discussed and interpreted. The significance of this research in the immediate context of the field of biostatistics and bioinformatics is examined. Recommendations for further research end the chapter. In summary, this chapter is divided into four sections: 5.1 Summary of results, 5.2 Conclusion, 5.3 Recommendations and 5.4 Suggestions for future studies.

#### 5.1 Summary of Results

Classification trees for competing information are helpful for arranging patients into several groups with comparable risks. We developed tree for Univariate competing risk. We proposed Within-node homogeneity by using Deviance residual and Cox-Snell residual as a response in a common classification tree and Between-node Heterogeneity using Gray (1988), Pepe and Mori (1991,1993), Kolmogorov-Smirnov test (2007, 2014) and Neyman Smooth test (2007) as splitting rule to compare the two CIFs present in the nodes. The methods particularly intended to model survival data with competing risks.

Under Within-node homogeneity, all the proposed models seem to fit the data structure better than the existing model in the exponential and lognormal distributions through the simulation schemes. In this context, the proposed method detects cut-point than the existing methods when the failure time follows an exponential distribution irrespective of the censoring rates and detects the cut point equally if the estimate is around 60%. In Log-normal failure time, the proposed model tends to behave equally with the existing method except when the cut point is more than 0.1 and less than or equal to 0.7, the tables of their simulation techniques and results were given in chapter three.

When the real-life data were considered, better results were observed from the extended models than the existing method. Additionally, results from the empirical studies via bone marrow transplant data showed that the proposed model with Cox-Snell residual performs better than both the Martingale residual and Deviance residual respectively.

Equally, results from Prostate cancer when Cox-Snell residual was considered, generated  $MSE=0.01783563$  and  $deviance=14.9732200$ , which is less than those produced by Deviance and Martingale residuals. This implies that, the proposed model performed better than the existing methods. These results validate those obtained from the Monte-Carlo studies.

Generally, when it comes to data structure performance, no methods appeared to be able to select covariate or number of terminal nodes than the proposed methods. In this case, Cox-Snell residual in within-node homogeneity classification tree performed better than any existing methods in terms of detecting the accurate cut-point and data structure for covariate associated to survival times.

Furthermore, when considering between-node heterogeneity, the estimated mean and median were used as the measures of indices based on simulation scheme, thus, we use the results to check the performance of our proposed models based on the cutpoint detection. Correspondingly, we observed that proposed models have a value that are much closer to the cutpoint estimate than the existing techniques at 50% censoring rates and 0.5 cutpoint. Consequently, the relative performances of the proposed models were checked by ranking the results in tables 4.13 and 4.16, all in the rank level of 8 based on the sample sizes, we observed that the Neyman smooth test has the highest relative performance of 50% and 75% of mean and median than any other methods, and figs. 4.23 & 4.24 affirm the results irrespective of the distribution.

Another method considered to check the performance of the extended methodology is the system time (seconds). This measures the efficiency of the test and their relative performance in term of the system time, the results were provided for both exponential and log-normal in Tables 4.15 & 4.18. We observed from the results that as the sample size increases the amount of time to run any of the selected tests increases. However, the extended method by Pepe test statistic was found to be highly effective in terms of system time while Neyman Smooth is best in terms of the median.

A real-life data set on follicular cell lymphoma with two covariates age and hemoglobin was used to test the performance of the proposed diagnostic classification tree based on between node heterogeneity. The results in Table 4.20 were ranked based on their closeness to the estimated mean and median, the results show that the proposed models perform better than the existing model. Further estimation was done on this real-life data by using the methods proposed by Contal and O'Quigley (1999) to check the optimum cutpoint of the data. Optimum cutpoint selected for the patient age is 57.4 years and hemoglobin level is 140g/l, and only age was found to influence ( $p\text{-value}<0.05$ ) the survival patient from the disease. Comparably, the proposed diagnostic classification tree estimate that is close to the optimum cutpoint 547.4 years is pepe with 52 years of age and Cox-PH with 139g/l hemoglobin level.

The performance of ANN was implemented in this work. The study extended the ANN to survival analysis with competing risks, and the tremendous progress of this new extension is very impressive and encouraging. Additionally, comparison between the proposed model and classical competing risk modeling was done. However, three different model performance assessment methods were employed, these are: root mean square error (RMSE), mean absolute error (MAE) and relative squared error (RSE) for the comparison. Moreover, three different real-life dataset were used through out this work and only double-blinded randomized clinical trial was interpreted

critically. Other dataset were employed to validate the performance of the proposed model and the existing model. Likewise, variable importance was performed on the double-blinded randomized clinical trial dataset through the use of Garson Algorithm for both competing risks.

All the analysis implemented on the proposed ANN was done through the use of package "survival", "KMsurv", "NeuralNetTools", "RSNNS", "nnet", "devtools", "neuralnet" and "clusterGeneration" all in the language R.

This work also examines the efficiency of a new technique for modeling survival MARS with Cox-Snell residuals relative to some of the existing methods (the Cox PH and Martingale residual methods) under three different distributions of survival times – Exponential, Weibull and Gompertz distributions.

Monte-Carlo experiment was carried out to examine the behaviors of some metrical covariate in the two-modelling structure in which the Cox-Snell and Martingale residuals were used as responses in the survival MARS models under low, medium and high censoring rates at different values of penalty (tuning) parameters.

Numerical results from the fitted survival MARS models with Cox-Snell and Martingale residuals on the simulated data were presented in Tables 4.26 (for Exponential survival times), 4.27 (for Weibull distributed survival times) and 4.28 (for Gompertz distributed survival times). Graphical summaries of the results in Tables 4.26, 4.27 and 4.28 were further presented in Appendices B (for Exponential survival times), C (for Weibull distributed survival times) and D (for Gompertz distributed survival times) respectively.

It was observed from the simulation results in Table 4.26 that for Exponentially distributed survival times, the proposed Cox-Snell residual model detected the nonlinear influence of  $x_1$  and

its interactions with  $x_2$  more than the Martingale residual (existing) model at the penalizing parameter  $d = 2$  and at low (25%) and medium (50%) censoring rates. However, the Martingale residual model detected the nonlinear influence of  $x_1$ , its interactions with  $x_2$  and the amount of false chosen basis functions (in %) at the high (75%) censoring rate more than the Cox-Snell residual model.

Further results from Table 4.26 showed that when the penalized parameter is 3 and the censoring rates were low, medium and high the Cox-Snell residual model detected the non-linear influence of  $x_1$ , its interaction with  $x_2$  and the amount of false chosen basis function than the Martingale residual model.

Without loss of generality, it was found that the efficiency of the Cox-Snell residual model increases in term of its ability to detect the non-linearity of  $x_1$  and its interaction with  $x_2$  as the value of the penalized parameter increases, although at high censoring rate.

Similar results were obtained from the survival MARS models with Cox-Snell and Martingale residuals for Weibull and Gompertz distributed survival times.

Finally, the proposed survival MARS model with Cox-Snell residuals, the survival MARS model with Martingale residual and the classical Cox PH models were applied on a real-life dataset to validate the results from Monte-Carlo studies. The results obtained showed that the proposed Cox-Snell residual model was more efficient than both the existing Martingale residual survival MARS and the classical Cox PH models.

## 5.2 Conclusion

This work has established the use of diagnostic classification trees based on within-node homogeneity, between-node heterogeneity, Cox-Snell Multilayers perceptron in artificial neural network on competing risk survival data and multivariate adaptive regression splines to the survival data without competing risks. By these extensions, an integrated method that can handle: medical domain knowledge, potential predicting factors on cancer diseases (candidate patients) and the prognostic risk of the diseases has been achieved. Further conclusions are made in following manners.

Firstly, the study extended within-node homogeneity by using Deviance residual and Cox-Snell residual as a response in a common classification tree. The proposed methods were good for detecting the accurate cutpoint for covariate associated to survival times and excellent in classifying the risk of disease than the existing methods based on the empirical simulation and real-life data.

Secondly, the between-node heterogeneity method was found to perform better in assigning subjects to groups with different levels of risk and detecting the optimum cutpoint than the existing methods. This was achieved based on the simulation results and real-life data, with the complement of some extended test statistics.

Thirdly, this study has successfully proposed artificial neural network in competing risk model, using resilient back-propagation. We applied three different datasets on proposed model and existing model, we found that the proposed model performs better than the existing one irrespective of the performance assessment methods. In addition, when the Garson's Algorithm was applied to the data set, it was observed that one of the independent variables

(Electrocardiogram Code) predicts the survival of the patients with the risk of dying from prostate cancer better than any other covariates. Moreover, age of the patients also predicts the survival of patients with prostate cancer when the risk is due to other diseases. Finally, Tumour Stage, Cardiovascular Disease History and Performance Rating were found to influence the survival of the patients with prostate cancer irrespective of the risk.

Finally, in this work, the use of (modified) Cox- Snell Residual to Survival analysis with MARS was proposed. The proposed method was compared with Martingale Residual Survival MARS mode ling technique and the classical Cox-PH model. Both the modified Cox-Snell and Martingale residuals of a Cox-PH model were used as responses in a common survival analysis with MARS.

All the results from simulation studies and real-life application reveal that the proposed survival MARS model with Cox-Snell residuals fitt the data more appropriately than either the Martingale residual survival MARS or the classical Cox-PH model.

### **5.3 Recommendations**

Based on the results of this study, the following recommendations are made:

- i. The classical methods of analyzing survival data do not gives satisfactory results at the local or regional levels when the underlying assumptions that govern the usability of these models failed and when the survival data is of high dimension. In this regard, the extended data mining techniques proposed in this study are hereby recommended to circumvent these challenges.
- ii. Researchers that wish to classify subjects/objects into sub-groups based on their levels of risk should use the extended classification tree.

- iii. For any study that involve modelling the linear, non-linear and the interaction effect of the covariates, the extended ANN and MARS are hereby recommended.

#### **5.4 Suggestions for Future Studies**

The following are the possible areas to be exploited in future research:

1. The proposed classification trees method in this study can be extended to model survival data with time-dependent covariates, i.e. situations where the assumption of proportional hazards is not fulfilled.
2. There is need to extend the proposed Artificial Neuron Network (ANN) into survival time data without competing risk using Cox-Snell residual.
3. The proposed MARS methods need to be extended to survival time data with competing risk endpoints.

## REFERENCES

- Aalen, O. and Johansen, S (1978). An empirical transition matrix for non-homogeneous Markov chains based uncensored observations. *Scandinavian Journal of Statistics*. 5:141–150.
- Andersen, P. K., Borgan, Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- Andrej Krenker, Janez Bester and Andrej Kos (2011). Introduction to the Artificial Neural Networks, *Artificial Neural Networks - Methodological Advances and Biomedical Applications*, Prof. Kenji Suzuki (Ed.), ISBN:978-953-307-243-2, InTech, Available from: <http://www.intechopen.com/books/artificial-neural-networks-methodological-advances-and-biomedical-applications/introduction-to-the-artificial-neural-networks>
- Bajorunaite R. and Klein J. P.(2008) Comparison of failure probabilities in the presence of competing risks. *J. Stat. Comput. Simul.*, 78:951–966.
- Bajorunaite, R. & Klein, J. P. (2007). Two-sample tests of the equality of two cumulative incidence functions. *Comput. Statist. Data Anal.* 51, 4269-4281.
- Bani K. M., David G., Denison T., Adrian F. and Smith M. (2000). Bayesian Survival analysis using a MARS Model.
- Ben-Ari E. N. and Steinberg D. M. (2007). Modeling data from computer experiments: An empirical comparison of kriging with mars and projection pursuit regression. *Quality Engineering*, vol. 19, pp. 327–338, 2007.
- Bender R., Augustin T. and Blettner M (2005): Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24:1713–1723.
- Beyersmann J., Pauly M., and Di Termini S (2013). Weak Convergence of the Wild Bootstrap for the Aalen-Johansen Estimator of the Cumulative Incidence Function of a Competing Risk. *Scandinavian Journal of Statistics*, 40:387–402.
- Breiman L., Friedman J., Olshen R., and Stone C., (1984). *Classification and Regression Trees*. Wadsworth, Belmont California.

- Breslow N. (1974). Covariance analysis of censored survival data, *Biometrics* 30, 89–99
- Bruce D. B. and Craig E. R. (1999). A comparison of conventional linear regression methods and neural networks for forecasting educational spending, *Economics of Education Review*, 18(4), pp.405-415.
- Burlak, G., Muñoz, J., Ochoa A. and Hernández J.A. (2006), Detecting Cheats In Online Student Assessments Using Data Mining, In International Conference on Data Mining, Las Vegas, pp. 204-210.
- Chap T. L. (2003): Introduction Biostatistics. John Wiley & Sons, Inc., Hoboken, New Jersey.
- Chipman, H., George, E., and McCulloch, R. (1998) “Bayesian CART Model Search (with discussion)”, *Journal of the American Statistical Association*, 93, 935–960.
- Cleves M. A., Gould W.W. and Gutierrez R.G (2004): An Introduction to Survival Analysis Using Stata. Stata Press, College Station, Texas.
- Contal C., O’Quigley J (1999). An application of change point methods in studying the effect of age on survival in breast cancer. *Computational Statistics & Data Analysis* **30**(3):253-70.
- Cortes C. and Vapnik V. (1995). Support-vector networks. *Machine Learning*. **20** (3): 273–297.
- Costa M. J. and Shaw J.E.H. (2006). Parametrization and Penalties in Spline Models with an Application to Survival Analysis. Paper No. 08-02, [www.warwick.ac.uk/go/crism](http://www.warwick.ac.uk/go/crism).
- Cox D.R. (1972): Regression models and life tables (with discussion). *Journal of the Royal Statistical Society - B* 1972; 34: 187-220.
- Crowder M.J. (2012): Multivariate Survival Analysis and Competing Risks. Chapman & Hall/CRC Press, Boca Raton, London New York.
- Crowder, M.J. (2001). Classical Competing Risks. Chapman & Hall/CRC, 1st edn.
- Dauda K.A, Yahya W.B and Banjoko A.W (2015): Survival Analysis With Multivariate Adaptive Regression Splines Using Cox-Snell Residual. *Anale. Seria Informatică*. 13(2): 25-41.

- David Kraus (2008). surv2sample: Two-Sample Tests for Survival Analysis. R package version 0.1-2. <http://www.davidkraus.net/surv2sample/>
- David, H.A. and Moeschberger, M.L.(1978).Theory of Competing Risks(Griffin's Statistical Monographs and Courses:No.39). Lubrecht & Cramer Ltd.
- Dobler D. and Pauly M. Pauly (2013). How to Bootstrap Aalen-Johansen Processes for Competing Risks? Handicaps, Solutions and Limitations. Submitted paper.
- Elith J. and Leathwick J. (2007): Predicting species distribution from museum and herbarium records using multi-response models fitted with multivariate adaptive regression splines, in: Diversity and Distribution, 13:265-275.
- Fayyad U.M (1996). Data Mining and Knowledge Discovery: Making Sense Out of Data, IEEE Expert 11(5): 20-25.
- Felici, G. and Vercellis, C. (2007). Mathematical Methods for Knowledge Discovery and Data Mining. Idea Group Reference.
- Feng Gao, Amita K. Manatunga, Shande Chen (2006). Developing Multivariate Survival Trees with a Proportional Hazards Structure. Journal of Data Science 4: 343-356.
- Fine J. P. and Gray R. J. (1999) A proportional hazards model for the sub-distribution of a competing risk. J Am Stat Assoc. 94:496–509.
- Fiona C. M. (2008). Classification Trees For Survival Data With Competing Risks. Department of Biostatistics, University of Pittsburgh (Unpublished).
- Frauke G. and Stefan F. (2010). Neuralnet: Training of Neural Networks. R Journal Vol. 2/1, June.
- Friedman J. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics 19: 1 - 141.
- Fritsch and Günther F. (2008). Neuralnet: Training of Neural Networks. R Foundation for Statistical Computing. R package version 1.2.

- Garret J. M. (1997) : Graphical Assessment of the Cox Model Proportional Hazard Assumption. Stata Technical Bulletin 35:9-14. In Stata Bulletin Reprint 6:38-44. College Station, TX: Stata Press.
- Garson G.D (1991). Interpreting Neural-Network Connection Weights. *AI Expert*; 6:46-51.
- Gray R. J (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Statist.*, 16:1141–1154.
- Gurney K. (1997). *An Introduction to Neural Networks*. (1<sup>st</sup> ed.) UCL Press, London EC4A 3DE, UK.
- Hastie T., and Tibshirani R (1990). *Generalized Additive Models* (1990), Chapman & Hall, New York.
- Hastie T., Tibshirani R. and Friedman J. H. (2001): *The Elements of Statistical Learning: Prediction, Inference and Data Mining*. Springer Verlag, New York.
- Hess K.R. (1995): Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression. *Statistics in Medicine* 14: 1707-1723.
- Hornik K., Stichcombe M., and White H. (1989). Multi-Layer Feedforward Networks Are Universal Approximations. *Neural Networks*, 2:359–366.
- Hwang, G.J., Tsai, P.S., Tsai, C.C. and Tseng, J.C.R. (2008). A novel approach for assisting teachers in analyzing student web-searching behaviors. In *Computer & Education Journal*, 51, pp. 926-938.
- Ibrahim N. A. and Kudus A. (2009). Decision Tree for Prognostic Classification of Multivariate Survival Data and Competing Risks, *Recent Advances in Technologies*, Maurizio A Strangio (Ed.), ISBN: 978-953-307-017-9.
- Ibrahim N. A., Kudus A., Daud I., and Abu Bakar M. R.(2008). Decision Tree for Competing Risks Survival Probability in Breast Cancer Study. *International Journal of Biological and Life Sciences*, 3:1.

- Ishwaran H., Gerds T.A., Kogalur U.B., Moore R.D., Gange S.J. and Lau B.M. (2014). Random survival forests for competing risks. *Biostatistics*, 15(4), 757-773.
- Jain A.K. and Mao J. (1994). Neural Networks and Pattern Recognition, in *Computational Intelligence: Imitating Life*, J.M. Zurada, R. J. Marks 11, and C.J. Robinson, eds., IEEE Press, Piscataway, N.J., pp. 194-212.
- Jonas K., Patrik E. and Mattias O. (2013). Ensembles of genetically trained artificial neural networks for survival analysis. *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges (Belgium), 24-26, i6doc.com publ., ISBN 978-2-87419-081-0.
- Kalderstam, J. (2015). *Neural Network Approaches To Survival Analysis* Department of Astronomy and Theoretical Physics, Lund University.
- Klein J. P and Moeschberger, M. L (2003). *Survival analysis: Techniques for censored and truncated data*. New York: Springer.
- Klein, J. P. and Moeschberger M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*, New York: Springer-Verlag.
- Kooperberg C. (2013): *Polspline: Polynomial spline routines*. R package version 1.1.9. <http://CRAN.R-project.org/package=polspline>.
- Kooperberg C., Bose S., and Stone C. J (1997). Polychotomous Regression. *Journal of the American Statistical Association* 92, Issue 437: 117 – 127.
- Kraus, D. (2007). Data-driven smooth tests of the proportional hazards assumption. *Lifetime Data Anal.* 1:31-16.
- Krenker A.; Volk M.; Sedlar U.; Bešter J.; Kos A. (2009). Bidirectional artificial neural networks for mobile-phone fraud detection. *ETRI Jurnal.*, V31:92-94, COBISS.SI-ID 6951764
- Kriner M. (2007): *Survival Analysis with Multivariate adaptive Regression Splines*

- Kuhnert P. M. and McClure R. (2000). Combining non-parametric models with logistic regression: an application to motor vehicle injury data. *Computational Statistics and Data Analysis*, vol. 34, pp. 371–386.
- LeBlanc M and Crowley J (1999). Adaptive Regression Splines in the Cox Model. *Biometrics* 55: 204 – 213.
- LeBlanc M. and J. Crowley (1992). Relative risk trees for censored survival data. *Biometrics*, 48:411–425.
- LeBlanc M. and J. Crowley (1993). Survival trees by goodness of split. *Journal of the American Statistical Association*, 88(422):457–467.
- Lin D. Y (1997). Non-parametric inference for cumulative incidence functions in competing risks studies. *Statistics and Medicine*, 16:901–910.
- Malgorzata K. and Leon B. (2000). Artificial Neural Networks in Identifying Areas with Homogeneous Survival Time.
- McCulloch W. and Walter P. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 5 (4): 115–133.
- Muehlenbrok M. (2005). Automatic action analysis in an interactive learning environment, In workshop on Usage Analysis in Learning Systems at the 12<sup>th</sup> International Conference on Artificial Intelligence in Education, Amsterdam TheNetherlands.73-80.2005.
- Orna I. and Charles K. (1995). Trees and splines in survival analysis. *Statistical Methods in Medical Research*, 5:237-261.
- Paula F. and Joao T. (2008). Applying the artificial neural network methodology for forecasting the tourism time series. 5<sup>th</sup> International conference Business and Management, 16-17, Vilnius, Lithuania.
- Pepe M. S (1991). Inference for events with dependent risks in multiple endpoint studies. *J.Amer. Statist. Assoc.*, 86(415):770–778.

- Pepe, M.S. & Mori, M. (1993) Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* 12: 737-751.
- Piatetsky S. G. and Frawley W.J. (1991). *Knowledge Discovery in Databases*. AAAI/MIT.
- Prentice R. L., Kalbfleisch J. D., Peterson A. V., J., Flournoy, N., Farewell, V. T., and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 34(4): 541–554.
- Psichogios D. C., De-Veaux R. D., and Ungar L. H. (1992). Non-parametric system identification: A comparison of mars and neural networks. *American Control Conference*, pp. 1436–1441.
- R Development Core Team- R (2015): A language and environment for statistical computing. R foundation for Statistical computing, Vienna, Austria, URL <http://www.r-project.org/>.*
- Rayner, J. C. W. & Best, D. J. (1989). *Smooth Tests of Goodness of Fit*. Oxford University Press, New York.
- Ripley B. D. (1996): *Pattern Recognition and Neural\_Networks*. Cambridge University Press, 7:403-521.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations* Cambridge, MA: MIT Press.
- Sima J.(1998). *Introduction to Neural Networks*, Technical Report No. V755, Institute of Computer Science, Academy of Sciences of the Czech Republic.
- Society B, 34(2), 187-220.
- Stein A. (2011): *Residuals and Functional Form in Accelerated Life Regression Models*. Norwegian University of Science and Technology Department of Mathematical Sciences, M.Sc. thesis.
- Su X. and Fan J.(2001). *Multivariate survival trees by goodness of split*. Technical Report 367. Department of Statistics, University of California, Davis.

Therneau T. M., Grambsch P.M. and Fleming T. R (1990). Martingale-based residuals for survival models, *Biometrika*, 77:147-160.

Therneau T.M., Grambsch P.M., and Fleming T (1990) Martingale based residuals for survival models. *Biometrika*, 77:147–160.

Tsai J.C.C. and Chen V.C.P. (2005): Flexible and robust implementations of multivariate adaptive regression splines within a wastewater treatment stochastic dynamic pro-gram, in: *Quality and Reliability Engineering International*, 21:689-699..

Tsai J.C.C., and Chen V.P. (2005: Flexible and robust implementations of multivariate adaptive regression splines within a wastewater treatment stochastic dynamic pro-gram. *Quality and Reliability Engineering International*, 21: 689-699.

Tsiatis, A. A. (1975). Aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 72: 20-22.

[URL:http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:amore.](http://wiki.r-project.org/rwiki/doku.php?id=packages:cran:amore)

Vander Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. With Applications Statistics*. Springer, New York.

Vapnik, V. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York.

Venables W. and Ripley B (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. URL: <http://www.stats.ox.ac.uk/pub/MASS4>. ISBN 0-387-95457-0.

Yahya W. B., Morolake O. O. and Emmanuel T. J. (2012). A Fast Algorithm to Construct Neural Network Classification Models with High-Dimensional Genomic Data. *Anal Seria Information*. Vol.Xfasc.39-56.

Waiyamai K. (2003). *Improving Quality of Graduate Students by Data Mining*. Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkok, Thailand.

Weber G.W., Batmaz I., Koksall G., Taylan P. and Yerlikaya-Ozkurt F. (2012): CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines

supported by continuous optimization, Inverse Problems in Science and Engineering (IPSE) 20:371-400.

Wei Fu and Jeffrey S. Simonoff (2016). Survival trees for left-truncated and right-censored data, with application to time-varying covariate data. arXiv:1606.03033v1 [stat.ME].

Witten, I. H. and Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques with JAVA Implementations (2nd Ed.). Morgan Kaufmann, San Francisco.

Yahya WB, Aremu GT and Garba MK (2015): Multiclass Sequential Feature Selection and Classification Method for Genomic Data. *Journal of Applied Science and Technology*, 20 (1&2): 50-60.

Yahya W. B. and Ulm K. (2009). Survival Analysis of Breast and Small-Cell Lung Cancer Patients using Conditional Logistic Regression Model. *Int'l. Jour. of Ecological Economics & Statistics*, 14, (S09): 15-34. <http://www.ceser.res.in/ijees.html>

Yahya W. B. and Dauda K. A. (2015): Explorative Study of Duration of Marriage In Nigeria: Survival Analysis Approach.

Yahya WB (2009): Sequential dimension reduction and prediction methods with high dimensional microarray data. Universitätsbibliothek, Ludwig- Maximilians-Universität, München, Germany. Ph.D. Thesis. URL: <http://edoc.ub.uni-muenchen.de/10254/>

Yahya WB (2012), Genes selection and tumour classification in cancer research: A new approach. Lambert Academic Publishing, Säbruck, Germany.

Yahya WB, Rosenberg R and Ulm K. (2014): Microarray-based Classification of Histopathologic Responses of Locally Advanced Rectal Carcinomas To Neoadjuvant Radiochemotherapy Treatment. *Turkiye Klinikleri Journal of Biostatistics*, 6(1):8-23.

Yahya WB, Ulm K, Ludwig F and Hapflemeyer A. (2011): K-SS: A sequential feature selection and prediction method in microarray study. *International Journal of artificial intelligence*, Spring, volume 6, number S11.

- Yan J. (2012): KMsurv Data sets from Klein and Moeschberger Survival Analysis. R package version 0.1-5. <http://CRAN.R-project.org/package=KMsurv>.
- Yanying Yang (2010). Neural Network Survival Analysis. Unpublished M.Sc. dissertation.
- Yuhong Wu, Tjelmeland H and Mike West.(2006) Bayesian CART-Prior Specification and Posterior Simulation.
- Zaïane, O. and Luo, J. (2001). Web usage mining for a better web -based learning environment. In Proceedings of Conference on Advanced Technology for Education. Banff, Alberta, pp. 60–64.
- Zhou Y., and Leung H. (2007): Predicting object-oriented software maintainability using multivariate adaptive regression splines. *Journal of Systems and Software*, 80: 1349-1361.
- Zhou Y., and Leung H. (2007): Predicting object-oriented software maintainability using multivariate adaptive regression splines, in: *Journal of Systems and Software*, 80:1349-1361.
- Žilinskas, A., Žilinskas, J. (2010). Interval arithmetic based optimization in nonlinear regression. *Informatika*, 21(1), 149–158.

## APPENDICES

### Appendix A:

Table A1: Results from the Contal and O'Quigley's method for the covariate Age

|      | agef | U          | Q           | p-value     |
|------|------|------------|-------------|-------------|
| 1:   | 57.4 | 30.4289893 | 1.866369665 | 0.001885572 |
| 2:   | 58.0 | 30.0359877 | 1.842264813 | 0.002254714 |
| 3:   | 57.2 | 29.9898290 | 1.839433654 | 0.002302211 |
| 4:   | 57.5 | 29.7420472 | 1.824235900 | 0.002573403 |
| 5:   | 57.6 | 29.4465194 | 1.806109629 | 0.002935385 |
| ---  |      |            |             |             |
| 239: | 84.3 | 0.2424559  | 0.014871091 | 1.000000000 |
| 240: | 84.0 | 0.2174441  | 0.013336989 | 1.000000000 |
| 241: | 85.0 | 0.1843777  | 0.011308852 | 1.000000000 |
| 242: | 82.7 | 0.1035367  | 0.006350449 | 1.000000000 |
| 243: | 17.0 | 0.0000000  | 0.000000000 | 1.000000000 |

Table A2: Results from the Contal and O'Quigley's method for the covariate Hemoglobin (g/l)

| S/N | hgb | U          | Q           | p-value   |
|-----|-----|------------|-------------|-----------|
| 1:  | 140 | 5.06130450 | 0.310436376 | 0.9999777 |
| 2:  | 146 | 4.86285744 | 0.298264576 | 0.9999920 |
| 3:  | 142 | 4.12257624 | 0.252859244 | 1.0000000 |
| 4:  | 147 | 3.69966202 | 0.226919694 | 1.0000000 |
| 5:  | 143 | 3.56077654 | 0.218401119 | 1.0000000 |
| 6:  | 129 | 3.40764695 | 0.209008877 | 1.0000000 |
| 7:  | 145 | 3.39091851 | 0.207982834 | 1.0000000 |
| 8:  | 135 | 3.18688478 | 0.195468374 | 1.0000000 |
| 9:  | 153 | 3.16263246 | 0.193980853 | 1.0000000 |

|     |     |            |             |           |
|-----|-----|------------|-------------|-----------|
| 10: | 141 | 3.07431742 | 0.188564028 | 1.0000000 |
| 11: | 151 | 3.03448515 | 0.186120905 | 1.0000000 |
| 12: | 127 | 2.91640731 | 0.178878571 | 1.0000000 |
| 13: | 126 | 2.79697052 | 0.171552886 | 1.0000000 |
| 14: | 125 | 2.70053668 | 0.165638092 | 1.0000000 |
| 15: | 130 | 2.29797300 | 0.140946748 | 1.0000000 |
| 16: | 148 | 2.29534154 | 0.140785347 | 1.0000000 |
| 17: | 144 | 2.12641676 | 0.130424304 | 1.0000000 |
| 18: | 114 | 1.99856902 | 0.122582730 | 1.0000000 |
| 19: | 165 | 1.99082669 | 0.122107853 | 1.0000000 |
| 20: | 128 | 1.98250869 | 0.121597666 | 1.0000000 |
| 21: | 156 | 1.92633813 | 0.118152431 | 1.0000000 |
| 22: | 152 | 1.83263016 | 0.112404829 | 1.0000000 |
| 23: | 166 | 1.70278860 | 0.104440964 | 1.0000000 |
| 24: | 120 | 1.59976505 | 0.098121989 | 1.0000000 |
| 25: | 118 | 1.55043574 | 0.095096363 | 1.0000000 |
| 26: | 154 | 1.46534202 | 0.089877119 | 1.0000000 |
| 27: | 164 | 1.46527133 | 0.089872783 | 1.0000000 |
| 28: | 116 | 1.45367499 | 0.089161519 | 1.0000000 |
| 29: | 139 | 1.41825085 | 0.086988770 | 1.0000000 |
| 30: | 161 | 1.30469065 | 0.080023527 | 1.0000000 |
| 31: | 113 | 1.29409682 | 0.079373752 | 1.0000000 |
| 32: | 172 | 1.22596983 | 0.075195166 | 1.0000000 |
| 33: | 110 | 1.04705265 | 0.064221236 | 1.0000000 |
| 34: | 136 | 1.02154107 | 0.062656477 | 1.0000000 |
| 35: | 149 | 1.01574241 | 0.062300814 | 1.0000000 |
| 36: | 160 | 0.98404307 | 0.060356528 | 1.0000000 |

|     |     |            |             |           |
|-----|-----|------------|-------------|-----------|
| 37: | 155 | 0.91817598 | 0.056316553 | 1.0000000 |
| 38: | 158 | 0.89863832 | 0.055118206 | 1.0000000 |
| 39: | 157 | 0.80111751 | 0.049136743 | 1.0000000 |
| 40: | 105 | 0.62021909 | 0.038041293 | 1.0000000 |
| 41: | 150 | 0.57718991 | 0.035402087 | 1.0000000 |
| 42: | 124 | 0.52639656 | 0.032286664 | 1.0000000 |
| 43: | 189 | 0.49229161 | 0.030194829 | 1.0000000 |
| 44: | 132 | 0.45167163 | 0.027703392 | 1.0000000 |
| 45: | 86  | 0.44200721 | 0.027110623 | 1.0000000 |
| 46: | 134 | 0.41792878 | 0.025633766 | 1.0000000 |
| 47: | 177 | 0.38078920 | 0.023355801 | 1.0000000 |
| 48: | 131 | 0.32082019 | 0.019677586 | 1.0000000 |
| 49: | 163 | 0.31167797 | 0.019116846 | 1.0000000 |
| 50: | 80  | 0.30157528 | 0.018497195 | 1.0000000 |
| 51: | 123 | 0.29137148 | 0.017871343 | 1.0000000 |
| 52: | 162 | 0.24164654 | 0.014821451 | 1.0000000 |
| 53: | 117 | 0.13395107 | 0.008215922 | 1.0000000 |
| 54: | 137 | 0.11036810 | 0.006769455 | 1.0000000 |
| 55: | 170 | 0.05561196 | 0.003410973 | 1.0000000 |
| 56: | 100 | 0.05028440 | 0.003084206 | 1.0000000 |
| 57: | 138 | 0.03525735 | 0.002162519 | 1.0000000 |
| 58: | 133 | 0.03190508 | 0.001956906 | 1.0000000 |
| 59: | 40  | 0.00000000 | 0.000000000 | 1.0000000 |

**Appendix B: Graphs of Survival MARS results for Simulated Exponential Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters.**

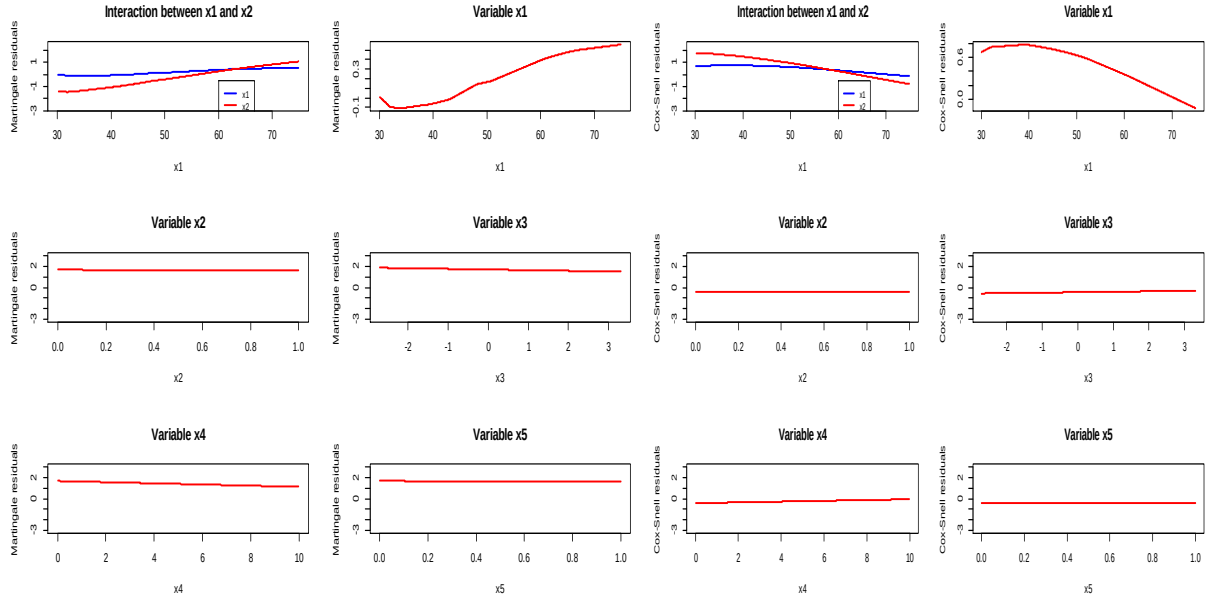


**Fig. 1:** Graphs of Survival MARS results under Exponential Survival times with penalized parameter  $d = 2$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

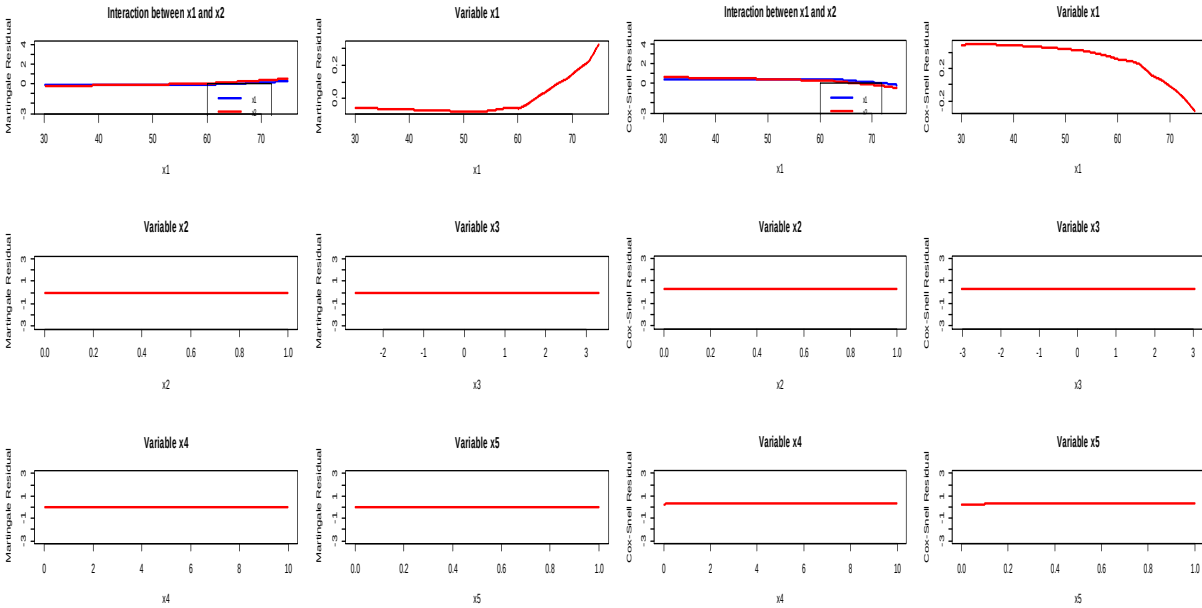


**Fig. 2:** Graphs of Survival MARS results under Exponential Survival times with penalized parameter  $d = 2$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

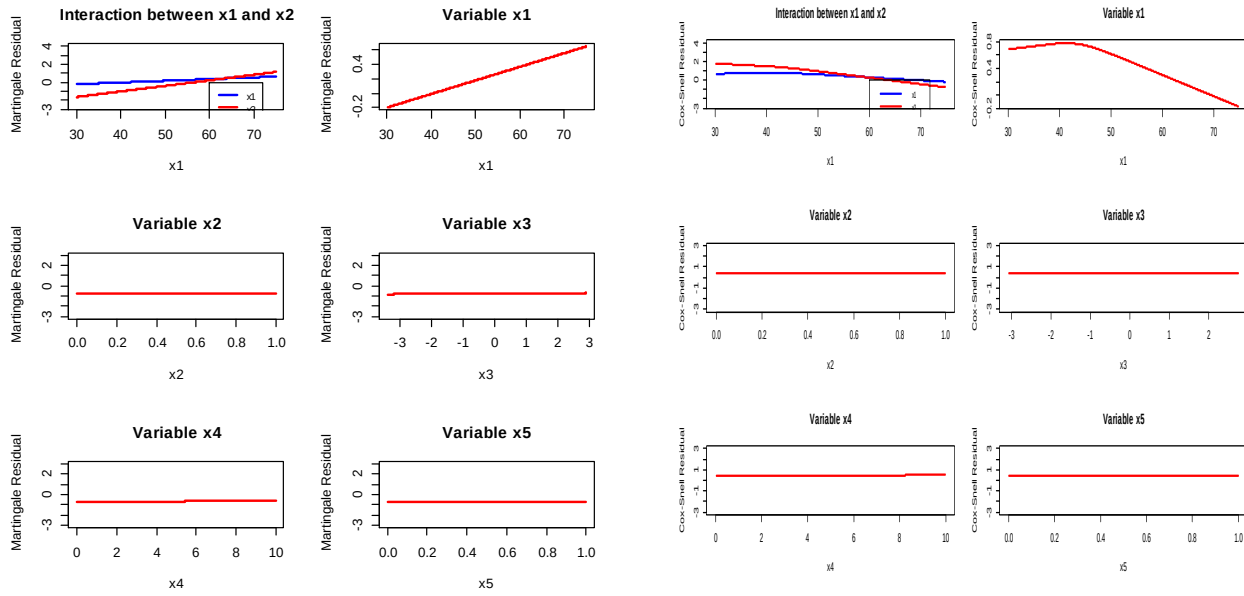


**Fig. 3:** Graphs of Survival MARS results under Exponential Survival times with penalized parameter  $d = 5$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

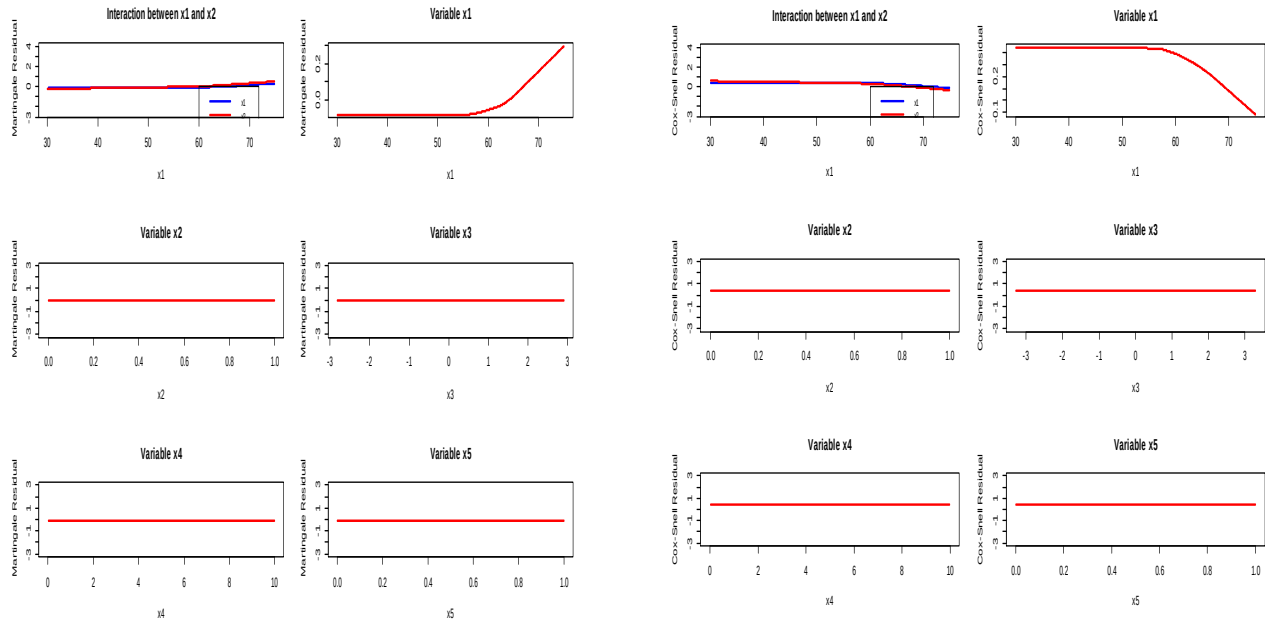


**Fig. 4:** Graphs of Survival MARS results under Exponential Survival times with penalized parameter  $d = 5$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

## Appendix C: Graphs of Survival MARS results for Simulated Weibull Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters.

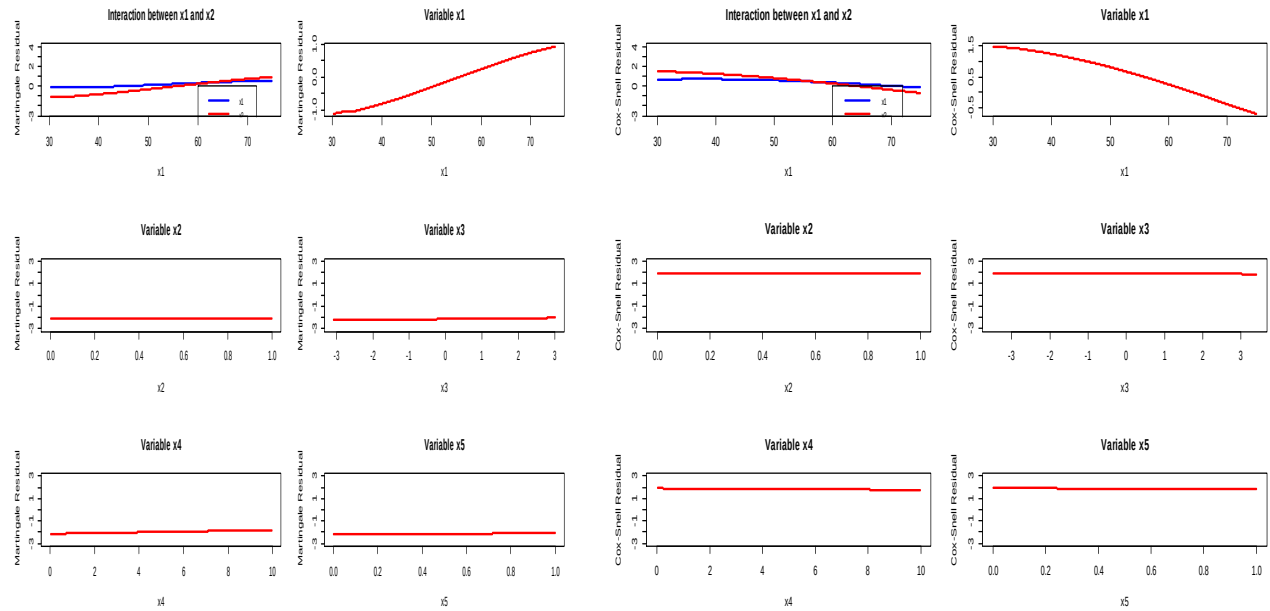


**Fig. 5:** Graphs of Survival MARS results under Weibull Survival times with penalized parameter  $d = 2$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

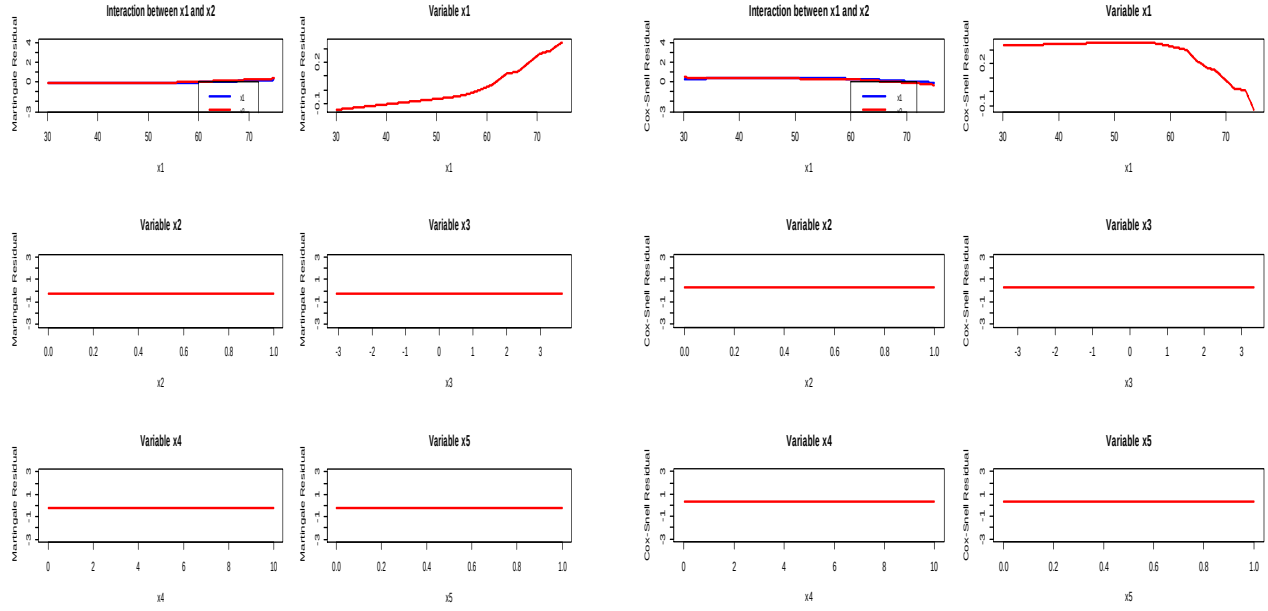


**Fig. 6:** Graphs of Survival MARS results under Weibull Survival times with penalized parameter  $d = 2$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

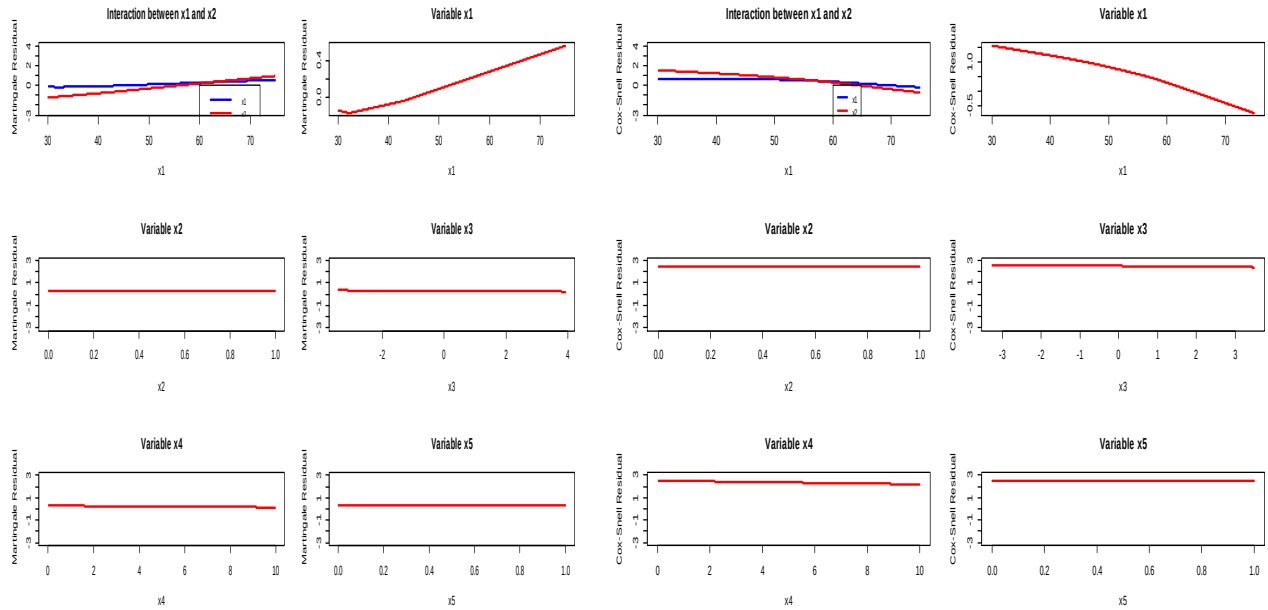


**Fig. 7:** Graphs of Survival MARS results under Weibull Survival times with penalized parameter  $d = 5$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

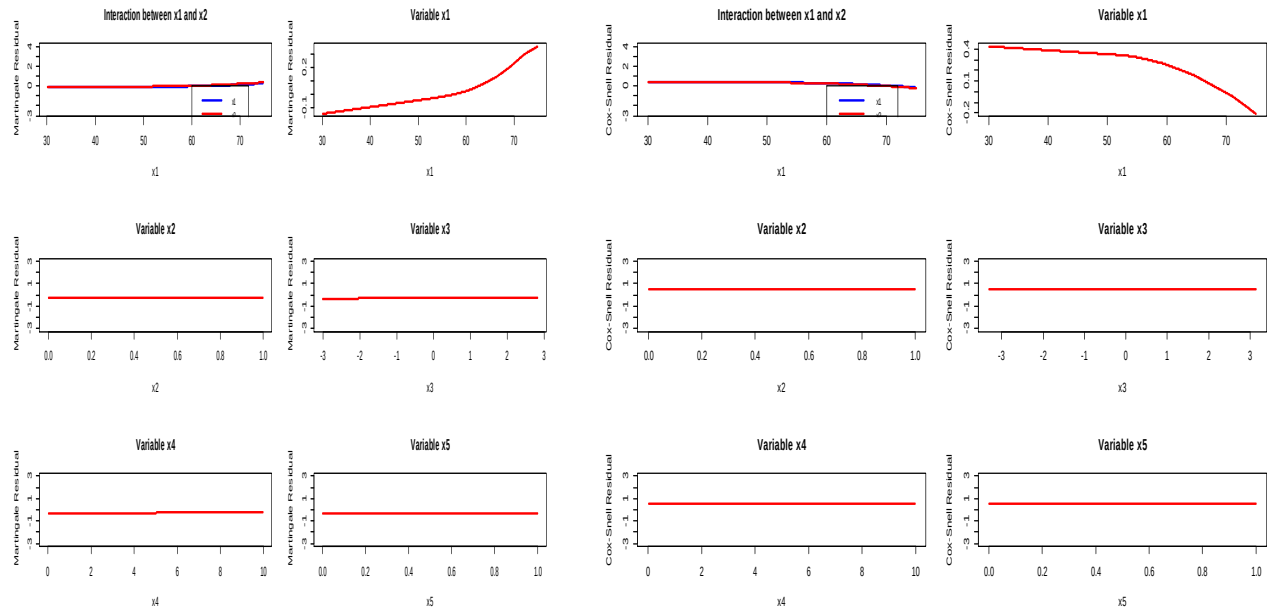


**Fig. 8:** Graphs of Survival MARS results under Weibull Survival times with penalized parameter  $d = 5$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

**Appendix D: Graphs of Survival MARS results for Simulated Gompertz Survival Times with Martingale and Cox-Snell Residuals at different penalty parameters.**

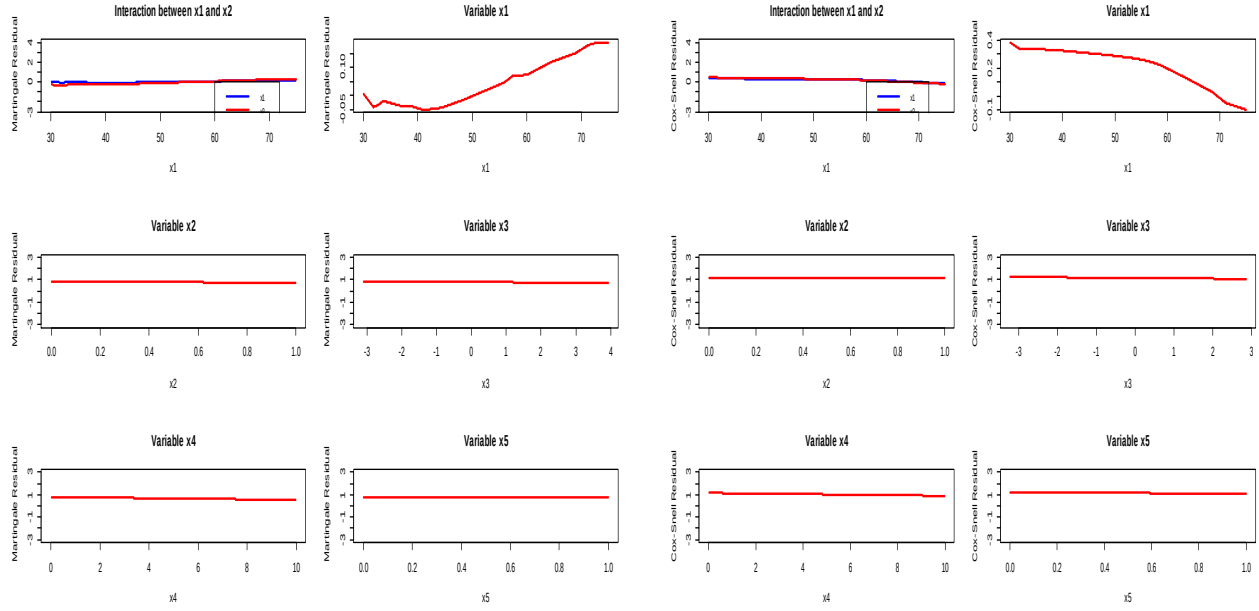


**Fig. 9:** Graphs of Survival MARS results under Gompertz Survival times with penalized parameter  $d = 2$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

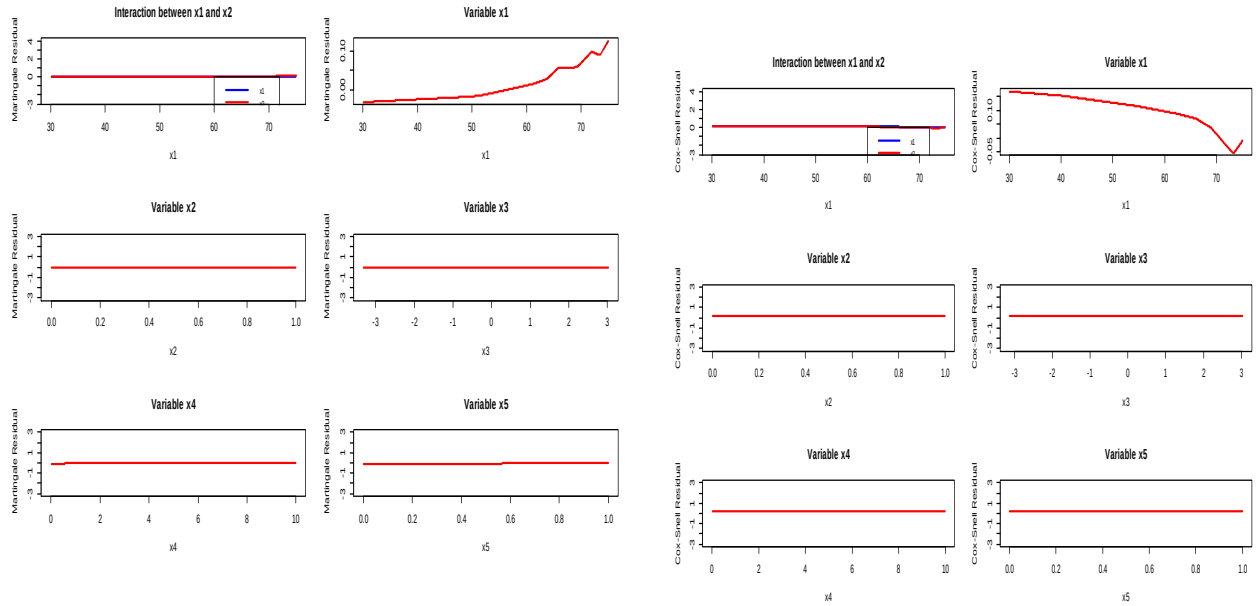


**Fig. 10:** Graphs of Survival MARS results under Gompertz Survival times with penalized parameter  $d = 2$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

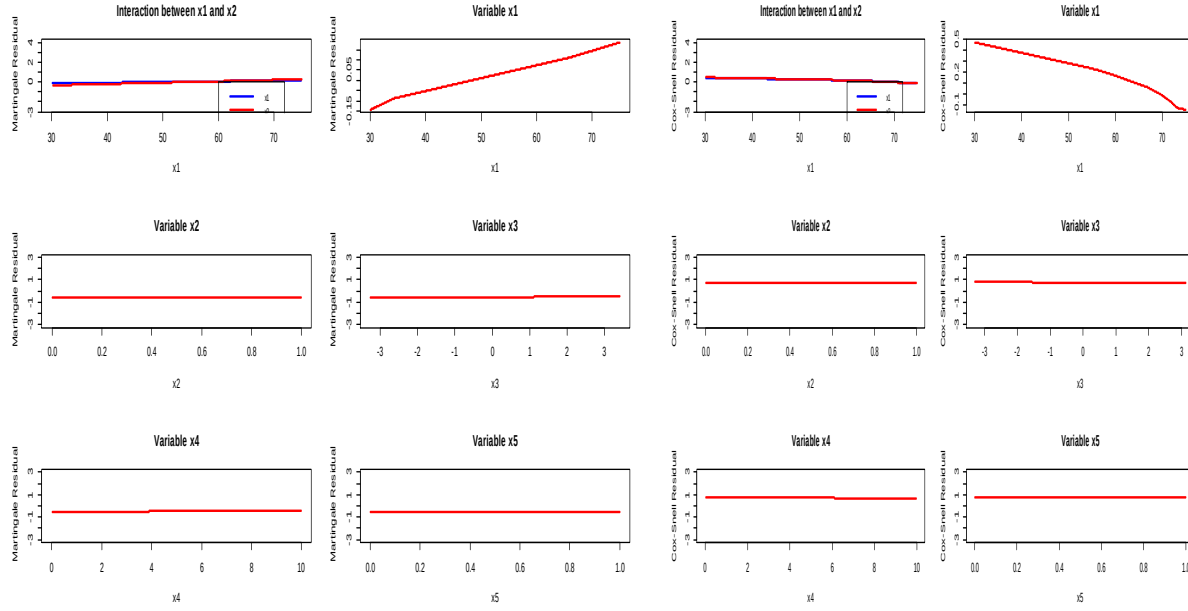


**Fig. 11:** Graphs of Survival MARS results under Gompertz Survival times with penalized parameter  $d = 5$  at low (25%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

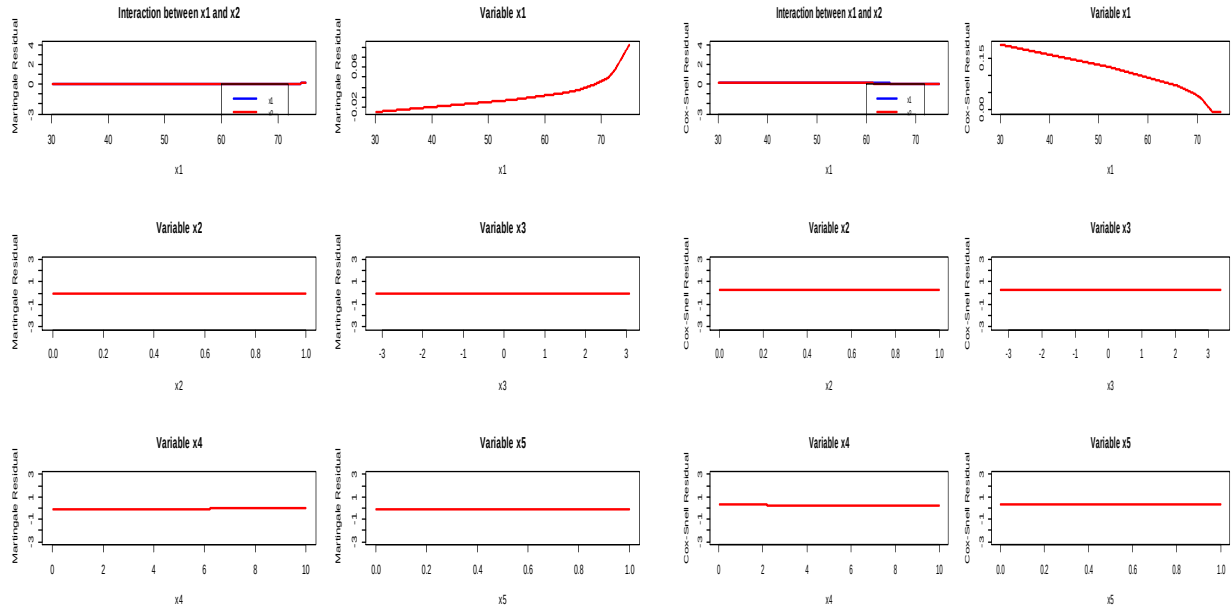


**Fig. 12:** Graphs of Survival MARS results

under Gompertz Survival times with penalized parameter  $d = 5$  at high (85%) censoring rate for Martingale (left graph) and Cox-Snell (right graph) Residuals.

## Appendix E: R Codes for both Simulation and Real Life Data

### E1: Classification Tree for within node homogeneity code

```
library(survival)
library(cmpsrsk)
library(rpart)
library(mvpart)
library("rpart")
library("evtree")
library(timereg)
library(randomForest)
```

#### Code E1.1: Cox-Snell Residual Simulation Based on Cutpoint

```
simcox<-function(c1,c2,r1,r2,r3,r4,r5,r6,r7,B,ss,ss2){
tempMR<-c();
tempMR2<-c();
nodevec<-rep(0,ss);
for(i in 1:B){
Z<-runif(ss);
t1<-rexp(ss,r1)*(Z<=c1);
124
t1<-t1+rexp(ss,r2)*(Z>c1);
t2<-rexp(ss,r3)*(Z<=c2);
t2<-t2+rexp(ss,r4)*(Z>c2);
t3<-rexp(ss,r5)*(Z<=c1);
t3<-t3+rexp(ss,r6)*(Z>c1&Z<=c2);
t3<-t3+rexp(ss,r7)*(Z>c2);
tsim<-pmin(t1,t2,t3);
dsim<-(t1<t2)*(t1<t3)+2*(t2<t1)*(t2<t3);
# gray stats are calculated
# the LR stats are calculated
dsim2<-(dsim==1)^2;
# the residual stats
temp2MR<-coxph(Surv(tsim,dsim2)~1);
martnull <- residuals(temp2MR,type="martingale");
coxsnell<-(dsim2-martnull)
tempMR<-c(tempMR,rpart(coxsnell~Z,method="anova",
maxdepth=1,minbucket=ss2)$splits[4])
tempMR2<-c(tempMR2,rpart(coxsnell~Z,method="anova",
maxdepth=1,cp=0,minbucket=ss2,dissim="man")$splits[4])
}
outputMR<-tempMR;
outputMR2<-tempMR2
output<-cbind(outputMR,outputMR2);
output
}
##### 25%, 50% and 75% censoring#####
#simcox<-simcox(0.5,0.3,.2,.4,.7,.56,.1,.24,.5,2,200,20)
```

#### Code E1.2: Deviance Residual Simulation Based on Cutpoint

```
simDev<-function(c1,c2,r1,r2,r3,r4,r5,r6,r7,B,ss,ss2){
```

```

tempMR<-c();
tempMR2<-c();
nodevec<-rep(0,ss);
for(i in 1:B){
Z<-runif(ss);
t1<-rexp(ss,r1)*(Z<=c1);
t1<-t1+rexp(ss,r2)*(Z>c1);
t2<-rexp(ss,r3)*(Z<=c2);
t2<-t2+rexp(ss,r4)*(Z>c2);
t3<-rexp(ss,r5)*(Z<=c2);
t3<-t3+rexp(ss,r6)*(Z>c2&Z<=c1);
t3<-t3+rexp(ss,r7)*(Z>c1);
tsim<-pmin(t1,t2,t3);
dsim<-(t1<t2)*(t1<t3)+2*(t2<t1)*(t2<t3);
dsim2<-(dsim==1)^2;
# the residual stats
temp2MR<-coxph(Surv(tsim,dsim2)~1);
devinull <- residuals(temp2MR,type="deviance")
tempMR<-c(tempMR,rpart(devinull~Z,method="anova",
maxdepth=1,minbucket=ss2)$splits[4])
tempMR2<-c(tempMR2,rpart(devinull~Z,method="anova",
maxdepth=1,cp=0,minbucket=ss2,dissim="man")$splits[4])
}
outputMR<-tempMR;
outputMR2<-tempMR2
output<-cbind(outputMR,outputMR2);
output
}
##### 25%, 50% and 75% censoring#####
#simDev<-simDev(0.25,0.3,.2,.4,.7,.56,.1,.24,.04,2,200,20)

```

### Code E1.3: Martingale Residual Simulation Based on Cutpoint

```

simMar<-function(c1,r1,r2,r3,r4,r5,B,ss,ss2){
tempMR<-c();
tempMR2<-c();
nodevec<-rep(0,ss);
for(i in 1:B){
Z<-runif(ss);
t1<-rexp(ss,r1)*(Z<=c1);
t1<-t1+rexp(ss,r2)*(Z>c1);
t2<-rexp(ss,r3)*(Z<=c1);
t2<-t2+rexp(ss,r4)*(Z>c1);
t3<-rexp(ss,r5);
tsim<-pmin(t1,t2,t3);
dsim<-(t1<t2)*(t1<t3)+2*(t2<t1)*(t2<t3);
dsim2<-(dsim==1)^2;
# the residual stats
temp2MR<-coxph(Surv(tsim,dsim2)~1);
martingale<- residuals(temp2MR,type="martingale")
tempMR<-c(tempMR,rpart(martingale~Z,method="anova",
maxdepth=1,minbucket=ss2)$splits[4])
tempMR2<-c(tempMR2,rpart(martingale~Z,method="anova",
maxdepth=1,cp=0,minbucket=ss2,dissim="man")$splits[4])
}
outputMR<-tempMR;
outputMR2<-tempMR2;
output<-cbind(outputMR, outputMR2);

```

```
output
}
```

```
##### 25%, 50% and 75% censoring#####
#simMar<-simMar(0.25,0.3,.2,.4,.7,.56,.1,.24,.04,2,200,20)
```

#### Code E1.4: R codes for the Tree Data Structures

```
set.seed(3435326)
simdata<-function(B,ss,r1,r2,r9, c1,c2,ss1,ss2,V, alphac){
MRstat<-c();
MRabsstat<-c();
DRstat<-c();
DRabsstat<-c();
CRstat<-c();
CRabsstat<-c();
for(i in 1:B){
#set.seed(2)
Z1<-runif(ss);
# discrete uniform (0,.25,.5,.75,1)
Z1<-floor(5*Z1)/4;
# Z2<-sample(c(0,1),ss,replace=T);
Z2<-runif(ss);
Z2<-floor(5*Z2)/4
t1<-rexp(ss,r1+r2*(Z1>c1& Z2>c2));
t2<-rexp(ss,1-r9-r1-r2*(Z1>c1& Z2>c2));
t3<-rexp(ss,r9);
# t1<-rexp(ss,r1);
# t2<-rexp(ss,r2+r3*(Z1>c1&Z2>c2));
# t3<-rexp(ss,r9);
tsim<-pmin(t1,t2,t3);
# vector with 0s 1s and 2s
dsim<-(t1<t2)*(t1<t3)+2*(t2<t1)*(t2<t3)
# the event indicator for LR test needs to be =1 if any event happens, and =0
ow.
#dsim2<-(dsim==1|dsim==2)^2
dsim2<-(dsim==1)^2;
Z3<-runif(ss);
Z3<-floor(5*Z3)/4;
Z4<-runif(ss);
Z4<-floor(5*Z4)/4;
# Z5<-sample(c(0,1),ss,replace=T);
# Z6<-sample(c(0,1),ss,replace=T);
Z5<-runif(ss);
Z5<-floor(5*Z5)/4;
Z6<-runif(ss);
Z6<-floor(5*Z6)/4;
nd<-rep(0,ss);
id<-seq(1:ss);
X<-data.frame(cbind(id,tsim,dsim,Z1,Z2,Z3,Z4,Z5,Z6,nd));
covariates<-cbind(Z1,Z2,Z3,Z4,Z5,Z6);
# MR
temp2MR<-coxph(Surv(tsim,dsim2)~1);
resids<-temp2MR$residuals;
tempMR<-rpart(resids~covariates,method="anova",minbucket=ss2);
tempMR2<-c(length(unique(tempMR$where)),covusedMR(tempMR,covariates));
```

```

# MR abs
# temp2MR<-coxph(Surv(tsim,d2sim)~1);
# resids<-temp2MR$residuals;
tempMRabs<-rpart(resids~covariates,method="anova",minbucket=ss2, dissim="man");
tempMR3<-c(length(unique(tempMRabs$where)),covusedMR(tempMRabs,covariates));
# DR
temp2DR<-coxph(Surv(tsim,dsim2)~1);
Davres <- residuals(temp2DR,type="deviance");
tempDR<-rpart(Davres~covariates,method="anova",minbucket=ss2);
tempDR2<-c(length(unique(tempDR$where)),covusedMR(tempDR,covariates));
# MR abs
# temp2MR<-coxph(Surv(tsim,d2sim)~1);
# resids<-temp2MR$residuals;
tempDRabs<-rpart(Davres~covariates,method="anova",minbucket=ss2, dissim="man");
tempDR3<-c(length(unique(tempDRabs$where)),covusedMR(tempDRabs,covariates));
# CR
temp2CR<-coxph(Surv(tsim,dsim2)~1);
Coxres <- dsim2-resids ;
tempCR<-rpart(Coxres~covariates,method="anova",minbucket=ss2);
tempCR2<-c(length(unique(tempCR$where)),covusedMR(tempCR,covariates));
# MR abs
# temp2MR<-coxph(Surv(tsim,d2sim)~1);
# resids<-temp2MR$residuals;
tempCRabs<-rpart(Coxres~covariates,method="anova",minbucket=ss2, dissim="man");
tempCR3<-c(length(unique(tempCRabs$where)),covusedMR(tempCRabs,covariates));
#second validation data set: currently without censoring
Z1<-runif(ss);
## discrete uniform (0,.25,.5,.75,1)
Z1<-floor(5*Z1)/4;
# Z2<-sample(c(0,1),ss,replace=T);
Z2<-runif(ss);
Z2<-floor(5*Z1)/4;
t12<-rexp(ss,r1+r2*(Z1>c1& Z2>c2));
t22<-rexp(ss,1-r9-r1-r2*(Z1>c1& Z2>c2));
# t3<-rexp(ss,r9);
# t12<-rexp(ss,r1);
# t22<-rexp(ss,r2+r3*(Z1>c1&Z2>c2));
# t32<-rexp(ss,r9);
# tsim2<-pmin(t12,t22,t32);
tsim2<-pmin(t12,t22);
## vector with 0s 1s and 2s
dsim2<-(t22<t12)+1
# dsim2<-(t12<t22)*(t12<t32)+2*(t22<t12)*(t22<t32)
d2sim2<-(dsim2==1)^2
# d2sim2<-(dsim2==1|dsim2==2)^2
# Z3<-floor(5*runif(ss))/4;
# Z4<-floor(5*runif(ss))/4;
Z3<-runif(ss);
Z3<-floor(5*Z3)/4;
Z4<-runif(ss);
Z4<-floor(5*Z4)/4;
Z5<-runif(ss);
Z5<-floor(5*Z5)/4;
Z6<-runif(ss);
Z6<-floor(5*Z6)/4;
covariates<-cbind(Z1,Z2,Z3,Z4,Z5,Z6);
nd2<-rep(0,ss);

```

```

id2<-seq(1:ss);
X2<-data.frame(cbind(id2,tsim2,dsim2,covariates,nd2));
####out
tempMR2<-c(tempMR2,maeMR(tempMR,X2,X,X$tsim,
X2$tsim2),maeMR2(tempMR,X2,X,X$tsim, X2$tsim2));
tempMR3<-c(tempMR3,maeMR(tempMRabs,X2,X,X$tsim,
X2$tsim2),maeMR2(tempMRabs,X2,X,X$tsim, X2$tsim2));
MRstat<-rbind(MRstat,tempMR2);
MRabsstat<-rbind(MRabsstat,tempMR3);
tempDR2<-c(tempDR2,maeMR(tempDR,X2,X,X$tsim,
X2$tsim2),maeMR2(tempDR,X2,X,X$tsim, X2$tsim2));
tempDR3<-c(tempDR3,maeMR(tempDRabs,X2,X,X$tsim,
X2$tsim2),maeMR2(tempDRabs,X2,X,X$tsim, X2$tsim2));
DRstat<-rbind(DRstat,tempDR2);
DRabsstat<-rbind(DRabsstat,tempDR3);
tempCR2<-c(tempCR2,maeMR(tempCR,X2,X,X$tsim,
X2$tsim2),maeMR2(tempCR,X2,X,X$tsim, X2$tsim2));
tempCR3<-c(tempCR3,maeMR(tempCRabs,X2,X,X$tsim,
X2$tsim2),maeMR2(tempCRabs,X2,X,X$tsim, X2$tsim2));
CRstat<-rbind(CRstat,tempCR2);
CRabsstat<-rbind(CRabsstat,tempCR3);
}
output<- rbind(MRstat,MRabsstat,DRstat,DRabsstat,CRstat,CRabsstat)
output
}
#aa= simdata (100,200,.1,.35,.5,.5,0,20,20,10,1)
#simdata<-data (25,500,.1,.35,.5,.5,0,20,20,10,1)
#write.table(simdata,file="simdata")

```

### Code E1.5: R codes for Real life Data

```

#####Real Life Data #####
library("rpart")
library("evtree")
library(survival)
library(timereg)
library(randomForest)
library(randomForestSRC)
#####
#####
##### cause 1
#####
bmt1<- read.csv(file = "C:\\Users\\Adesina\\Documents\\bmt1.csv")
attach(bmt1)
View(bmt1)
BMMR1<-coxph(Surv(time1,causel)~1)
BMMR1res<- residuals(BMMR1,type="martingale")
### CART
BMMRTree1<-rpart(BMMR1res~platelet1+age1 +tcell1,minbucket = 20,maxdepth = 20)
BMMRTree1
BMMRfor1<-randomForest(BMMR1res~platelet1+age1 +tcell1,minbucket = 20,maxdepth
= 20)
BMMRfor1
varImpPlot(BMMRfor1, main="",col="dark blue")
rp2DR <- as.party(BMMRTree1)
rp2DR

```

```

plot(rp2DR)
BMdat1=data.frame(BMMR1res,platelet1,age1,tcell11)
mtcars.obj <- rfsrc(BMMR1res ~., data = BMdat1)
plot(mtcars.obj,main="Variable Important plot for cause 1")
#####
####
BMDR1res<- residuals(BMMR1,type="deviance")
### CART
BMDRTree1<-rpart(BMDR1res~platelet1+age1 +tcell11,minbucket = 20,maxdepth = 20)
BMDRTree1
BMDRfor1<-randomForest(BMDR1res~platelet1+age1 +tcell11,minbucket = 20,maxdepth
= 20)
BMDRfor1
varImpPlot(BMDRfor1, main="",col="dark blue")
rp2DR <- as.party(BMDRTree1)
rp2DR
plot(rp2DR)
BDdat1=data.frame(BMDR1res,platelet1,age1,tcell11)
mtcars.DR <- rfsrc(BMDR1res ~., data = BDdat1)
plot(mtcars.DR,main="Variable Important plot for cause 1")
#### Cox-snell
BMCR1res=abs(cause1-BMMR1res)
BMCRTree1<-rpart(BMCR1res~platelet1+age1 +tcell11,minbucket = 20,maxdepth = 20)
BMCRTree1
BMCRfor1<-randomForest(BMCR1res~platelet1+age1 +tcell11,minbucket = 20,maxdepth
= 20)
BMCRfor1
varImpPlot(BMDRfor1, main="",col="dark blue")
rp2CR <- as.party(BMCRTree1)
rp2CR
plot(rp2CR)
BCdat1=data.frame(BMCR1res,platelet1,age1,tcell11)
mtcars.CR <- rfsrc(BMCR1res ~., data = BCdat1)
plot(mtcars.CR,main="Variable Important plot for cause 1")
##### Cause
2#####
bmt2<- read.csv(file = "C:\\Users\\Adesina\\Documents\\bmt2.csv")
attach(bmt2)
View(bmt2)
BMMR2<-coxph(Surv(time2,cause2)~1)
BMMR2res<- residuals(BMMR2,type="martingale")
### CART
BMMRTree2<-rpart(BMMR2res~platelet2+age2+tcell12,minbucket = 20,maxdepth = 20)
BMMRTree2
BMMRfor2<-randomForest(BMMR2res~platelet2+age2+tcell12,minbucket = 20,maxdepth =
20)
BMMRfor2
varImpPlot(BMMRfor2, main="",col="dark blue")
rp2DR2 <- as.party(BMMRTree2)
rp2DR2
plot(rp2DR2)
BMdat2=data.frame(BMMR2res,platelet2,age2,tcell12)
mtcars.obj2 <- rfsrc(BMMR2res ~., data = BMdat2)
plot(mtcars.obj2,main="Variable Important plot for cause 1")
#####
####
BMDR2res<- residuals(BMMR2,type="deviance")

```

```

### CART
BMDRtree2<-rpart(BMDR2res~platelet2+age2+tcell12,minbucket = 20,maxdepth = 20)
BMDRtree2
BMDRfor2<-randomForest(BMDR2res~platelet2+age2+tcell12,minbucket = 20,maxdepth =
20)
BMDRfor2
varImpPlot(BMDRfor1, main="", col="dark blue")
rp2DR2 <- as.party(BMDRtree2)
rp2DR2
plot(rp2DR2,main="CART due to Deviance residual through Cause 2")
BDdat2=data.frame(BMDR2res,platelet2,age2,tcell12)
mtcars.DR2 <- rfsrc(BMDR2res ~., data = BDdat2)
plot(mtcars.DR2,main="Variable Important plot for cause 2")
#### Modified Cox-snell
BMCR2res=abs(cause1-BMMR2res)
BMCRtree2<-rpart(BMCR2res~platelet2+age2+tcell12,minbucket = 20,maxdepth = 20)
BMCRtree2
BMCRfor2<-randomForest(BMCR2res~platelet2+age2 +tcell12,minbucket = 20,maxdepth
= 20)
BMCRfor2
varImpPlot(BMDRfor1, main="", col="dark blue")
rp2CR2 <- as.party(BMCRtree2)
rp2CR2
par(mfrow=c(2,1))
plot(rp2CR2,main="CART due to Modified Cox-Snell residual through Cause 2")
BCdat2=data.frame(BMCR2res,platelet2,age2,tcell12)
mtcars.CR2 <- rfsrc(BMCR2res ~., data = BCdat2)
plot(mtcars.CR2,main="Variable Important plot for cause 2")
#####
#####
##### Prostate cancer
#####
#####
#####
#####
#####
prostate <- read.csv(file = "C:\\Users\\Adesina\\Documents\\prostate.csv")
View(prostate)
attach(prostate)
##### Event 1#####
#####
##### Martingale #####
PMRT1<-coxph(Surv(TIME,event1)~1)
PMRT1res<- residuals(PMRT1,type="martingale")
PCMRtree1=rpart(PMRT1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCMRtree1
PCMRfor1<-
randomForest(PMRT1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCMRfor1
PC2MR <- as.party(PCMRtree1)
PC2MR
plot(PC2MR,main="CART due to Martingale residual through Cause 1 (Prostate
Cancer)")
PCdat1=data.frame(PMRT1res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)

```

```

PCMR.obj <- rfsrc(PMRT1res ~., data = PCdat1)
plot(PC.obj,main="Variable Important plot for cause 1")

##### Deviance Proposed #####
PMDR1res<- residuals(PMRT1,type="deviance")
### CART
PCDRTree1=rpart(PMDR1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCDRTree1
PCDRfor1<-
randomForest(PMDR1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCDRfor1
PC2DR <- as.party(PCDRTree1)
PC2DR
plot(PC2DR,main="CART due to Deviance residual through Cause 1 (Prostrate
Cancer)")
PCdat1=data.frame(PMDR1res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)
PCDR.obj <- rfsrc(PMDR1res ~., data = PCdat1)
plot(PCDR.obj,main="Variable Important plot for cause 1")
##### Cox_Snell Proposed #####
MRT1<-coxph(Surv(TIME,event1)~1)
PMRT1<- residuals(MRT1,type="martingale")
PMCR1res=abs(event1-PMRT1)
#PMDR1res<- residuals(PMRT1,type="deviance")
### CART
PCCRTree1=rpart(PMCR1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCCRTree1
PCCRfor1<-
randomForest(PMCR1res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCCRfor1
PC2CR <- as.party(PCCRTree1)
PC2CR
plot(PC2CR,main="CART due to Modifield Cox-Snell residual through Cause 1
(Prostrate Cancer)")
PCdat1=data.frame(PMCR1res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)
PCCR.obj <- rfsrc(PMCR1res ~., data = PCdat1)
plot(PCCR.obj,main="Variable Important plot for cause 1")
##### Cause 2 #####
##### Martingale #####
PMRT2<-coxph(Surv(TIME,event2)~1)
PMRT2res<- residuals(PMRT2,type="martingale")
PCMRTree2=rpart(PMRT2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCMRTree2
PCMRfor2<-
randomForest(PMRT2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCMRfor2
PC2MR2 <- as.party(PCMRTree2)
PC2MR2

```

```

plot(PC2MR2,main="CART due to Martingale residual through Cause 2 (Prostate
Cancer)")
PCdat2=data.frame (PMRT2res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)
PCMR.obj2 <- rfsrc(PMRT2res ~., data = PCdat2)
plot(PCMR.obj2,main="Variable Important plot for cause 1")

##### Deviance Proposed #####
PMDR2res<- residuals (PMRT2,type="deviance")
### CART
PCDRTree2=rpart (PMDR2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCDRTree2
PCDRfor2<-
randomForest (PMDR2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCDRfor2
PC2DR2 <- as.party(PCDRTree2)
PC2DR2
plot(PC2DR2,main="CART due to Deviance residual through Cause 2 (Prostate
Cancer)")
PCdat2=data.frame (PMDR2res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)
PCDR.obj2 <- rfsrc(PMDR2res ~., data = PCdat2)
plot(PCDR.obj2,main="Variable Important plot for cause 1")
##### Cox_Snell Proposed #####
MRT2<-coxph (Surv (TIME,event2)~1)
PMRT2<- residuals (MRT2,type="martingale")
PMCR2res=abs (event2-PMRT2)
#PMDR1res<- residuals (PMRT1,type="deviance")
### CART
PCCRTree2=rpart (PMCR2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG
, minbucket = 20,maxdepth = 20)
PCCRTree2
PCCRfor2<-
randomForest (PMCR2res~AGE+WTINDEX+PERF+HISTCD+ECGCODE+SIZEPT+SHINDEX+SERUMHG,
minbucket = 20,maxdepth = 20)
PCCRfor2
PC2CR2 <- as.party(PCCRTree2)
PC2CR2
plot(PC2CR2,main="CART due to Modifield Cox-Snell residual through Cause 2
(Prostate Cancer)")
PCdat2=data.frame (PMCR2res,AGE,WTINDEX,PERF,HISTCD,ECGCODE,SIZEPT,SHINDEX,SERUM
HG)
PCCR.obj2 <- rfsrc(PMCR2res ~., data = PCdat2)
plot(PCCR.obj2,main="Variable Important plot for cause 1")
##### Kidney transplant #####
samp=c (50,100,150,200,350,500,750,1000,2000)
NS=c (0.069,0.101,0.124,0.139,0.191,0.283,0.385,0.488,1)
KS=c (0.118,0.112,0.143,0.133,0.151,0.164,0.177,0.213,0.654)
pepe=c (0.077,0.084,0.109,0.116,0.144,0.155,0.206,0.249,0.667)
G=c (0.077,0.093,0.117,0.147,0.203,0.277,0.392,0.501,1)
Fine=c (0.069,0.097,0.118,0.15,0.206,0.283,0.406,0.519,1)
##### plot

```

```

plot(samp,NS,col=3,type="o",lwd=3,main="Power of the Five Selected Test Through
Cause 1",xlab="Sample Sizes",ylab="Power of the
Test",cex.lab=1.3,ylim=c(min(NS),max(NS)))
lines(samp,KS,col=4,lwd=3,type="o")
lines(samp,pepe,col=5,lwd=3,type="o")
lines(samp,G,col=2,lwd=3,type="o")
lines(samp,Fine,col=6,lwd=3,type="o")
legend("topleft",c("Neyman Smooth ","Kolmogorov Smirnov","Pepe
Integral","Gray","Fine and Gray"),lwd=3,lty=c(7),col=c(3,4,5,2,6))
##### cause 2
samp=c(50,100,150,200,350,500,750,1000,2000)
NS=c(0.043,0.077,0.059,0.076,0.056,0.055,0.065,0.067,0.242)
KS=c(0.067,0.078,0.083,0.069,0.06,0.073,0.075,0.06,0.253)
pepe=c(0.057,0.049,0.046,0.05,0.046,0.053,0.049,0.048,0.19)
G=c(0.05,0.064,0.043,0.059,0.056,0.052,0.057,0.057,0.205)
Fine=c(0.033,0.054,0.042,0.059,0.057,0.055,0.058,0.058,0.204)
##### plot
plot(samp,NS,col=3,type="b",lwd=3,main="Power of the Five Selected Test Through
Cause 2",xlab="Sample Sizes",ylab="Power of the
Test",cex.lab=1.3,ylim=c(min(NS),max(KS)))
lines(samp,KS,col=4,lwd=3,type="o")
lines(samp,pepe,col=5,lwd=3,type="o")
lines(samp,G,col=2,lwd=3,type="o")
lines(samp,Fine,col=6,lwd=3,type="o")
legend("topleft",c("Neyman Smooth ","Kolmogorov Smirnov","Pepe
Integral","Gray","Fine and Gray"),lwd=3,lty=c(7),col=c(3,4,5,2,6))
##### cutpoint real life data
#####
#####
prostate <- read.csv(file = "C:\\Users\\Adesina\\Documents\\prostate.csv")
View(prostate)
attach(prostate)
##### Event 1#####
#####
##### Martingale
#####
PMRT1<-coxph(Surv(TIME,event1)~1)
PMRT1res<- residuals(PMRT1,type="martingale")
PCMR1tree1=rpart(PMRT1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCMR1tree1
##### Deviance
#####
PMDR1res<- residuals(PMRT1,type="deviance")
### CART
PCDR1tree1=rpart(PMDR1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCDR1tree1
##### Cox_Snell Proposed #####
MRT1<-coxph(Surv(TIME,event1)~1)
PMRT1<- residuals(MRT1,type="martingale")
PMCR1res=abs(event1-PMRT1)
PCCR1tree1=rpart(PMCR1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCCR1tree1
##### Summary Statistics
event1#####
prostate3 <- read.csv(file = "C:\\Users\\Adesina\\Documents\\postrate 3.csv")
attach(prostate3)

```

```
View(prostate3)
```

```
##### Event 2#####
#####
##### Martingale
#####
PMRT1<-coxph(Surv(TIME,event2)~1)
PMRT1res<- residuals(PMRT1,type="martingale")
PCMR1tree1=rpart(PMRT1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCMR1tree1
##### Deviance
#####
PMDR1res<- residuals(PMRT1,type="deviance")
### CART
PCDR1tree1=rpart(PMDR1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCDR1tree1
##### Cox_Snell Proposed #####
MRT1<-coxph(Surv(TIME,event2)~1)
PMRT1<- residuals(MRT1,type="martingale")
PMCR1res=abs(event1-PMRT1)
PCCR1tree1=rpart(PMCR1res~AGE+WTINDEX+SERUMHG, minbucket = 20,maxdepth = 20)
PCCR1tree1
##### Summary Statistics
event1#####
prostate3 <- read.csv(file = "C:\\Users\\Adesina\\Documents\\cancer.csv")
attach(prostate3)
View(prostate3)
##### Residual plot #####
#####
### Plot of Martingale, Deviance and Cox Snell residuals against Age,WTINDEX
and SERUMHG
library(ggplot2)
par(mfrow=c(1,3))
qplot(AGE,PMRT1res, geom=c("point","smooth"),xlab="Patient
Age",ylab="Martingale Residuals", main="Figure 7: Plot of martingale residuals
against Age")
qplot(AGE,PMDR1res, geom=c("point","smooth"),xlab="Patient Age",ylab="Deviance
Residuals", main="Figure 8: Plot of deviance residuals against Age")
qplot(AGE,PMCR1res, geom=c("point","smooth"),xlab="Patient Age",ylab="Cox Snell
Residuals", main="Figure 9: Plot of cox snell residuals against Age")
##### Lowess
lo <- loess(PMRT1res~AGE)
plot(AGE,PMRT1res)
x1 <- seq(min(AGE),max(AGE), (max(AGE) - min(AGE))/1000)
lines(x1, predict(lo,x1), col='red', lwd=2)
```